2809076931

REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2006 Name of Author Li, Simon Yau-Wai

COPYRIGHT
This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION
I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOAN
Theses may not be lent to individuals, but the University Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: The Theses Section, University of London Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION
University of London theses may not be reproduced without explicit written permission from the University of London Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The University Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

☐ This copy has been deposited in the Library of UCL

☐ This copy has been deposited in the University of London Library, Senate House, Malet Street, London WC1E 7HU.

Bound By
Blissett Bookbind
020 8992 3965

An Empirical Investigation of Post-Completion Error: A Cognitive Perspective

Simon Yau-Wai Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of the
University of London

Department of Psychology
University College London

September 2006

UMI Number: U593061

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U593061
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I, *Simon Yau-Wai Li*, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Forgetting to retrieve your original after photocopying, forgetting to collect your card after a withdrawal from a cash machine, are examples of a specific type of omission error termed post-completion error (Byrne & Bovair, 1997). A post-completion error (PCE) is the omission of a "clean-up" step *after* the main goal of a task is fulfilled. The error phenomenon has the property of being infrequent but persistent; it does not occur very often and, yet, it continues to occur now and again.

This thesis is an empirical investigation of PCE to examine factors that provoke or mitigate the error. The investigation consists of two series of experiments. The first series of experiments is an extension of Byrne & Bovair's finding of the effect of high working memory demand on the increased occurrences of PCE. A novel paradigm was designed and adopted in the experiments; it was found that PCE also occurs in problem-solving tasks, which impose a high demand on working memory load. Results from the experiments also suggest that the use of static visual cues may reduce the error rate.

The second series of experiments investigates the effect of interruption on PCE in a procedural task paradigm. Based on the activation-based goal memory model (Altmann & Trafton, 2002) predictions were made on the effect of interruption position and duration on the error. Results show that PCE is more likely to occur with interruption occurring *just before* the post-completion step. Interruption occurring earlier in the task has no effect on PCE rate; it was found to be the same as having no interruptions at all. Moreover, interruption as brief as 15 seconds was found to be disruptive enough to increase PCE rate. The same disruptive effect was also obtained for other non-PCEs.

The scarcity and disparate nature of the existing theoretical approaches to PCE motivated a meta-theoretical analysis of PCE. The analysis has resulted in the identification of the major criteria required for an adequate account of PCE. Although a complete cognitive model of PCE is beyond the scope of the current thesis, the meta-theoretical analysis offers new insights into the understanding of PCE and aids future theoretical development.

The current thesis constitutes a methodological advance in studying PCE. New factors that provoke or mitigate the occurrence of the error were identified through empirical investigations. New insights into the understanding of the error were also possible through a meta-theoretical analysis within a coherent theoretical structure.

Acknowledgements

I would like to thank my supervisors, Prof. Ann Blandford, for her patient guidance throughout the PhD process, especially during my first year of study; Dr. Paul Cairns, for his constant enthusiasm and helpfulness. I am particularly grateful to Prof. Richard Young who unconditionally offers his supervision from which I am always inspired by his attitude and passion for academic research. I would also like to thank Dr. Anna Cox for her constant encouragement and occasional sarcastic comments, which I have always found funny when I manage to work them out!

I would like to acknowledge my friends at UCLIC and the Psychology department with whom I have enjoyed intellectual conversations, social support and alcohol consumptions with (I'm sure you can work out which group you belong to!). You guys ROCK! I would also like to thank Ricky Wong, Yong-Woo Lee and Martin Greaves for always being supportive and helpful. I am extremely pleased that I have met Magda Osman whose good nature, sarcasm and great taste for films have kept me going since the beginning of my study. Ok Magda, you are AMAZING!

I would like to express my gratitude to Prof. David Shanks and Dr. Alastair McClelland in the Psychology department for invaluable comments and discussion on my work. Big "thank you" also to Prof. Harold Thimbleby for caring support when I first joined UCLIC; to Prof. John Long for his inspiring comments and charming wit; to Prof. Wayne Gray for hosting part of my study in his lab at Rensselaer Polytechnic Institute, New York; and to Dr. Mike Scholles for his patience in teaching me Lisp programming.

During my stay at RPI in the States, I am pleased to have met a bunch of really fun people. I am especially grateful to Chris Myers, Tiffany Divine, Sean Li and Sun Xian for being caring and accommodating during my stay.

I would also like to acknowledge my friends outside of university: Ken Chan, Leon Chan, Jacker Chan, Carmen Choy, Jo Jo Fox, Ching Lau, Horace Lau, Jennifer Lau, Joanna Li, Loretta Li, Suki Mok, Geoff Poon, Kevin Sung, Alan "IC" Tang, Min Yan Wong, Danny Yip, and for being great friends, band mates, agony aunts and flatmates (again I'll let you work out which group(s) you belong to)!

Last, this acknowledgement will not be possible without my Dad, Mum, and sister who tirelessly offer me comfort throughout my entire PhD study.

Table of contents

ABSTRACT **3**

ACKNOWLEDGEMENTS **4**

TABLE OF CONTENTS **5**

LIST OF FIGURES **10**

LIST OF TABLES **11**

PUBLICATIONS **14**

CHAPTER 1: INTRODUCTION **15**

CHAPTER 2: LITERATURE REVIEW OF HUMAN ERROR RESEARCH **21**

1 INTRODUCTION 22
2 DIFFERENT PERSPECTIVES IN HUMAN ERROR RESEARCH 23
2.1 A broad perspective *23*
2.2 A narrow perspective *24*
3 CLASSIFICATION OF HUMAN ERROR IN VARIOUS TAXONOMIES 25
3.1 Phenomenological taxonomy *26*
3.2 Skill-based, rule-based and knowledge-based human performance taxonomy *26*
3.3 Distinction between slips and mistakes *28*
4 EXTANT THEORETICAL APPROACHES TO PCE 32
4.1 Supergoal kill-off approach *32*
4.2 Soar approach *33*
4.3 Activation-based goal memory approach *34*
4.4 Working memory capacity approach *36*
5 EMPIRICAL STUDIES OF HUMAN ERROR 38
5.1 Two empirical approaches to human error *38*
6 EMPIRICAL STUDIES OF PCE 41
7 SUMMARY 48

CHAPTER 3: POST-COMPLETION ERROR IN PROBLEM-SOLVING **50**

OVERVIEW 51

EXPERIMENT 3A: A PRELIMINARY STUDY **53**

1 INTRODUCTION 53
2 METHOD 53
2.1 Tasks *53*
2.2 Solving logic problems "in the head" using a text-based interface *54*
2.3 Design *56*
2.4 Materials *56*
2.5 Participants *57*
2.6 Procedure *57*
2.7 Measure *58*
3 RESULTS 58
3.1 Are the observed PCEs systematic in occurrence? *58*
3.2 Issues related to the obtained PCEs *58*
4 DISCUSSION 60

EXPERIMENT 3B: REFINED METHODOLOGY AND THE EFFECT OF STATIC VISUAL CUE **63**

1 INTRODUCTION 63
2 SOLVING LOGIC PROBLEMS "IN THE HEAD" 64
2.1 Two interfaces — text-based and menu-based *64*
3 METHOD 68
3.1 Tasks *68*
3.2 Design *69*
3.3 Materials *69*

3.4 Participants ... *70*
3.5 Procedure ... *70*
3.6 Measures ... *73*
4 RESULTS ... 73
4.1 Responses from checklist in the training trial ... *73*
4.2 PCE rates in Txt and Pop1 ... *74*
4.3 Difference in PCE occurrences between Txt and Pop1 condition ... *74*
4.4 PCE occurrences in each individual problem task ... *75*
4.5 Difference in number of PCEs between the three problem tasks ... *76*
5 DISCUSSION ... 76
5.1 Analysis on the three PC problems (DHC, Trch & IS) ... *77*
5.2 Pop-up menus in the Pop1 condition ... *82*

EXPERIMENT 3C: A FOLLOW-UP STUDY ... 84

1 INTRODUCTION ... 84
1.1 MODIFICATION OF THE POP-UP MENUS ... 84
2 METHOD ... 85
2.1 Tasks ... *85*
2.2 Design ... *85*
2.3 Materials ... *85*
2.4 Participants ... *85*
2.5 Procedure ... *86*
2.6 Measures ... *86*
3 RESULTS ... 86
3.1 Responses from checklist ... *86*
3.2 PCE rates in Txt and Pop2 ... *86*
3.3 Difference in PCE occurrences between conditions ... *87*
3.4 PCE occurrences in each individual problem task ... *87*
3.5 Difference in number of PCEs between the three problem tasks in Pop2 ... *88*
4 DISCUSSION ... 88
4.1 Limited effectiveness of static visual cues in mitigating PCEs ... *92*

OVERALL CHAPTER DISCUSSION ... 94

THEORETICAL DISCUSSION ... 97

CHAPTER 4: A META-THEORETICAL ANALYSIS OF POST-COMPLETION ERROR ... 102

1 INTRODUCTION ... 103
2 BASIS OF THE ANALYTICAL FRAMEWORK ... 104
3 ANALYSIS AND EVALUATION OF EXTANT PCE ACCOUNTS ... 106
3.1 Supergoal kill-off account ... *106*
3.2 Soar account ... *109*
3.3 Limited working memory capacity account ... *111*
3.4 Activation-based goal memory account ... *115*
4 DISCUSSION ... 120
4.1 Compare and contrast of the extant PCE accounts ... *120*
4.2 What can we learn from the current analysis? ... *124*
5 CONCLUDING REMARKS ... 125

CHAPTER 5: LITERATURE REVIEW OF INTERRUPTION RESEARCH ... 127

1 INTRODUCTION ... 128
2 EARLY BASIC PSYCHOLOGICAL RESEARCH ON INTERRUPTION ... 130
3 RECENT PSYCHOLOGICAL RESEARCH RELATED TO INTERRUPTION ... 131
3.1 Task switching ... *131*
3.2 Delayed-execute prospective memory ... *134*
4 SOME RECENT STUDIES ON TASK SIMILARITY AND COMPLEXITY ... 136
5 A COMMON THEORETICAL FRAMEWORK IN INTERRUPTION RESEARCH ... 139
6 REHEARSAL STRATEGY IN INTERRUPTION ... 140
7 THE EFFECT OF CUE AVAILABILITY ON INTERRUPTION ... 141
8 DETECTING THE DISRUPTIVE EFFECT OF AN INTERRUPTION USING MORE SENSITIVE MEASURES ... 143
9 THE EFFECT OF INTERRUPTION DURATION ... 144
10 THE EFFECT OF INTERRUPTION POSITION ... 146
11 SUMMARY AND THE RELATIONS BETWEEN INTERRUPTIONS AND PCE ... 148

CHAPTER 6: THE EFFECT OF INTERRUPTION ON POST-COMPLETION ERROR 152
OVERVIEW 153
EXPERIMENT 6A: A PRELIMINARY STUDY 155
1 INTRODUCTION 155
2 METHOD 156
2.1 Tasks *156*
2.2 Design *162*
2.3 Materials *163*
2.4 Participants *163*
2.5 Procedure *164*
2.6 Measures *164*
3 RESULTS 165
3.1 The effect of interruption on PCE rate *165*
3.2 Differential effects between the two types of interruptions *165*
3.3 The effect of interruption on task resumption times *167*
4 DISCUSSION 170
4.1 An analytical comparison and contrast between the doughnut-making task and Byrne & Bovair's Phaser task *172*
EXPERIMENT 6B: THE EFFECT OF INTERRUPTION POSITION 176
1 INTRODUCTION 176
2 METHOD 177
2.1 Tasks *177*
2.2 Design *182*
2.3 Materials *182*
2.4 Participants *183*
2.5 Procedure *183*
2.6 Measures *183*
3 RESULTS 184
3.1 Overall errors *184*
3.2 Interruption position effect on PCE *185*
3.3 Interruption position effect on non-PCE *186*
3.4 PCE and non-PCE resumption pattern *189*
3.5 Resumption times between different task steps *189*
3.6 Categorisation of non-PCE *191*
3.7 Interruption position effect on selector-step errors *193*
3.8 Errors at the Dough Port selector step *196*
4 DISCUSSION 196
EXPERIMENT 6C: THE EFFECT OF INTERRUPTION DURATION 200
1 INTRODUCTION 200
2 METHOD 201
2.1 Design *201*
2.2 Participants *201*
2.3 Materials and procedures *201*
2.4 Measures *201*
3 RESULTS 201
3.1 Overall errors *202*
3.2 Interruption position effect on PCE *203*
3.3 PCE and non-PCE resumption pattern *204*
3.4 Resumption times between different task steps *204*
3.5 Interruption position effect on non-PCE *206*
3.6 Categorisation of non-PCE *207*
3.7 Interruption position effect on selector-step errors *208*
3.8 Errors at "Selector DP" step *211*
4 DISCUSSION 211
EXPERIMENT 6D: A CHANGE TO THE PC STEP 215
1 INTRODUCTION 215
2 METHOD 216
2.1 Tasks *216*

2.2 Design *217*
2.3 Materials *218*
2.4 Participants *218*
2.5 Procedure *218*
2.6 Measures *219*
3 RESULTS 219
3.1 ANALYSIS OF PC TRIALS 219
3.1.1 Overall errors *219*
3.1.2 Interruption position effect on PCE *220*
3.1.3 PCE and non-PCE resumption pattern *221*
3.1.4 Resumption times between different task steps *222*
3.1.5 Interruption position effect on non-PCE *223*
3.1.6 Categorisation of non-PCE *224*
3.1.7 Interruption position effect on selector-step errors *224*
3.1.8 Errors at "Selector DP" step *227*
3.2 ANALYSIS OF NON-PC TRIALS 227
3.2.1 Overall errors *227*
3.2.2 Interruption position éffect on non-PCE *229*
3.2.3 Categorisation of non-PCE *230*
3.2.4 Interruption position effect on selector-step errors *231*
3.2.5 Errors at "Selector DP" step *233*
4 DISCUSSION 233

OVERALL CHAPTER DISCUSSION 236

THEORETICAL DISCUSSION 243

CHAPTER 7: CONCLUSION 248

FACTORS PROVOKING PCE 249
FACTOR MITIGATING PCE 250
Relations of PCE in problem-solving and procedural tasks *251*
THEORETICAL INSIGHTS INTO PCE 252
RELATION OF PCE TO OTHER ERRORS 253
EXPLAINING TWO REAL-WORLD PCE EXAMPLES 255
CONTRIBUTION TO KNOWLEDGE 256

REFERENCES 258

APPENDIX A (I) 266

INTERFACE INSTRUCTION MANUAL 266

APPENDIX A (II) 269

TRAINING PROBLEM SHEET 269

APPENDIX A (III) 270

RECALL SHEET 270

APPENDIX A (IV) 271

TRAINING PROBLEM INSTRUCTION SHEET 271

APPENDIX A(V) 272

REMINDER SHEET 272

APPENDIX A (VI) 273

THE FATHER & SON (FS) PROBLEM 273

APPENDIX A (VII) 274

THE DOG, HEN & CORN (DHC) PROBLEM 274

APPENDIX A (VIII) 275

THE SIMPLIFIED MISSIONARIES & CANNIBALS (SMC) PROBLEM 275

APPENDIX A (IX) 276

SOLUTIONS FOR THE FS PROBLEM: 276
SOLUTIONS FOR THE DHC PROBLEM: 276
SOLUTION FOR THE sMC PROBLEM: 277

APPENDIX B (I) 278

TRAINING PROBLEM SHEET 278

APPENDIX B (II) 279

CHECK LIST 279

APPENDIX B (III) 281

THE LION, MONKEY & BANANA (LMB) PROBLEM 281

APPENDIX B (IV) 282

THE 3 GUESTS (3G) PROBLEM 282

APPENDIX B (V) 283

THE TORCH (TRCH) PROBLEM 283

APPENDIX B (VI) 284

THE ITCHY & SCRATCHY (IS) PROBLEM 284

APPENDIX B (VII) 285

SOLUTION FOR THE TRAINING PROBLEM: 285

APPENDIX B (VIII) 286

SOLUTION FOR THE LMB PROBLEM: 286

APPENDIX B (IX) 287

SOLUTION FOR THE 3G PROBLEM: 287

APPENDIX B (X) 288

SOLUTION FOR THE TRCH PROBLEM: 288

APPENDIX B (XI) 289

SOLUTION FOR THE IS PROBLEM: 289

APPENDIX C 290

IS PROBLEM SPACE GRAPH 290
DHC PROBLEM SPACE GRAPH 291
TRCH PROBLEM SPACE GRAPH 292

APPENDIX D 293

APPENDIX E 297

FITTS' LAW CALCULATIONS FOR EXPERIMENT 6A 297
FITTS' LAW CALCULATIONS FOR EXPERIMENT 6B, 6C AND 6D 298

List of figures

Figure 3.1: A screen shot of the `main window` for entering solution steps of a puzzle 55

Figure 3.2: A screen shot of the `All steps` function being accessed at a particular point in time when solving the Missionary & Cannibal problem 56

Figure 3.3: A screenshot of the Txt interface 65

Figure 3.4: A screenshot of the History-All window 66

Figure 3.5: A screenshot of the Reminder window 66

Figure 3.6: A screenshot of the Pop1 interface 67

Figure 3.7: A screenshot of the object pop-up menu in the Pop1 interface 67

Figure 3.8: A screenshot of the river-side pop-up menu in the Pop1 interface 68

Figure 3.9: Pop-up menus used in Pop1 condition for DHC (left), Trch (middle) and IS (right 82

Figure 3.10: Modified pop-up menus for DHC (left), Trch (middle) and IS (right 84

Figure 6.1: The Wicket Doughnut Machine 157

Figure 6.2: Details of the steps required to operate the Doughnut Machine 158-159

Figure 6.3: A completion report when the task is completed 160

Figure 6.4: The Doughnut Packaging Express 161

Figure 6.5: The simple perceptual-motor interrupting task 162

Figure 6.6: A screen shot of the Tactical task (adapted, with permission, from Chung & Byrne (2004) which is the same as the Phaser task in Byrne & Bovair, 1997 174

Figure 6.7: The Wicket Doughnut Machine and the Call Centre 178

Figure 6.8: The Wicket Doughnut Machine 180

Figure 6.9: Details of the steps required to operate the Doughnut Machine 181-182

Figure 6.10: Error rate (%) of each task step across trials and participants 185

Figure 6.11: Error rate of each task step 202

Figure 6.12: Screen transitions of the beginning of a PC trial 216

Figure 6.13: Error rate of each task step in PC trials 220

Figure 6.14: Error rate of each task step in non-PC trials 228

List of tables

Table 3.1: Number of participants who recalled the transport vessel or included it in the problem solution 59

Table 3.2: The sentences used in indicating a post-completion step in the DHC, Trch and IS tasks 69

Table 3.3: Summary of number of PCEs obtained in the two experimental conditions 74

Table 3.4: The means, medians and interquartile ranges for the two conditions 74

Table 3.5: No. of PCEs in each individual problem task in condition Txt and Pop1 75

Table 3.6: Mean completion time and standard deviation (in bracket) for each PC problem task in the two experimental conditions 79

Table 3.7: Mean and standard deviation (in bracket) of History functions usage for each PC problem task in the two conditions 79

Table 3.8: The three problems (DHC, IS and Trch) and their corresponding characteristics in terms of four chosen dimensions 80

Table 3.9: Summary of number of PCEs obtained in the two experimental conditions 87

Table 3.10: The means, medians and interquartile ranges for the three conditions 87

Table 3.11: Individual problem task's PCE rate in Txt and Pop2 88

Table 3.12: The no. of PC problem tasks solved correctly in the two experimental Conditions 90

Table 3.13: Distribution of number of participants solving the PC problems correctly or incorrectly with PCE or no PCE in all three conditions 91

Table 3.14: *p* values of Fisher's Exact tests for assessing associations between PCEs and correct solutions for each problem task in three conditions 92

Table 6.1: GM and IR (in bracket) of the transformed resumption time data according to PC step and non-PC step 168

Table 6.2: GM and IR (in bracket) of the transformed resumption time data with movement times taken out according to different interruption positions 169

Table 6.3: A simple descriptive comparison of Byrne & Bovair's task and the current task in terms of task structure 173

Table 6.4: Number of PCEs and their mean error rates with respect to the different interruption positions 186

Table 6.5: Number of non-PCEs and their mean error rates with respect to the different interruption positions 188

Table 6.6: Mean resumption times of PC step and non-PC step 190

Table 6.7: Mean resumption times (minus estimated movement times) after interruption Z and Other ... 191

Table 6.8: The relative proportion (%) of the different identified error kinds ... 193

Table 6.9: Number and mean error rates of skip-selector errors according to different interruption position ... 194

Table 6.10: The number and mean error rates of incorrect-sequence error and PCE according to the different interruption positions ... 195

Table 6.11: Number of PCEs and their error rates with respect to different interruption positions in the two experimental conditions ... 203

Table 6.12: Mean resumption times (standard deviation in bracket) after interruptions Z and Other ... 205

Table 6.13: Mean resumption times with estimated movement times removed (standard deviation in bracket) after interruption Z and Other ... 205

Table 6.14: Number of non-PCEs and their mean error rates with respect to the different interruption positions ... 207

Table 6.15: The number and relative proportion (%) of the different error kinds in condition Short and Long ... 208

Table 6.16: Number and mean error rates of skip-selector errors according to different interruption position in condition Short and Long ... 209

Table 6.17: The number and mean error rates of incorrect-sequence error according to the different interruption positions in the two experimental conditions ... 210

Table 6.18: The distribution of interruption positions among the 18 trials in one testing session ... 217

Table 6.19: Number of PCE and their mean error rates with respect to the different interruption positions ... 221

Table 6.20: Mean resumption times after interruption Z and Other ... 222

Table 6.21: Mean resumption times (minus estimated movement times) after interruption Z and Other ... 222

Table 6.22: Number of non-PCEs and their mean error rates with respect to the different interruption positions ... 223

Table 6.23: The relative proportion (%) of the different identified error kinds ... 224

Table 6.24: Number and mean error rates of skip-selector errors according to different interruption position ... 225

Table 6.25: The number and mean error rates of incorrect-sequence errors and PCE according to the different interruption positions ... 226

Table 6.26: Number of xClean and their mean error rates with respect to the different interruption positions ... 229

Table 6.27: Number of non-PCEs and their mean error rates with respect to the different interruption positions........230

Table 6.28: The relative proportion (%) of the different identified error kinds........231

Table 6.29: Number and mean error rates of skip-selector errors according to different interruption position........231

Table 6.30: The number and mean error rates of incorrect-sequence errors and PCE according to the different interruption positions........232

Publications

Part of Experiment 3b has been presented as a poster and published as:

Li, S. Y. W., Blandford, A., Cairns, P., & Young, R. M. (2005). Post-completion errors in problem solving. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Conference.*

Experiment 6b and 6c have been presented as an oral presentation and published as:

Li, S. Y. W., Cox, A. L., Blandford, A., Cairns, P., Young, R. M., & Abeles, A. (2006). Further investigations into post-completion error: the effects of interruption position and duration. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Conference.*

Chapter 1

Introduction

"To err is human" sums up very succinctly that making errors is an inherent characteristic of human behaviour. The study of human error is not only important for practical reasons in the safety critical domains but is also interesting from a psychological perspective. Researchers such as Rasmussen (1982) and Reason (1990) suggest that rather than viewing human error as "fault", error can be viewed as an important part of behaviour that enables humans to learn and adapt to the environment.

To describe the particular kind of error that this thesis investigates, it is best to begin by giving some anecdotes from our everyday activity:

- Walking away from a photocopier with the copies in hand but leaving the original document behind.
- After withdrawing money from an automatic teller machine (ATM), walking away with the cash in hand but leaving the cash card behind.
- Having taken out dinner from an oven after 30 minutes of cooking time, forgetting to switch the oven off.
- Leaving change behind in a vending machine having bought the desired item.

The above anecdotes are all errors that share some common features:

- Each involves an action sequence that opens with satisfying some precondition (that the original be on the glass of the photocopier, that the card be inserted into the ATM, that the oven is switched on, that money is inserted into the vending machine).
- There is a subsequent action sequence that achieves the main goal of the interaction (that copies are obtained, that cash is acquired, that dinner is cooked, that change is collected).

- There is a final 'clean-up' step to restore the state of the system (original retrieved, card retrieved, oven switched off, change collected).

This type of omission error that involves omitting a functionally isolated step has been documented back in the 1980s (Rasmussen, 1980; 1982; 1987) and in the Human-Computer Interaction (HCI) literature as an error due to closure of a task (Dix, Finlay, Abowd & Beale, 1998). Curzon and Blandford (2001) provide a useful concept, termed interaction invariant, to describe the error. An interaction invariant describes that some part of a system to be perturbed in order to achieve the desired goal; the perturbations need to be undone at the end of the interaction. For example, an original document has to be placed under the lid of a photocopier in order to make the photocopies required, and the original has to be retrieved to return the photocopier to its initial state.

The term "post-completion error" (PCE) was introduced by Byrne and Bovair (1997) to describe the error phenomenon. Byrne and Bovair (1997, p. 32) define PCE as errors associated with a task characteristic that "some action is required to be performed *after* the main goal of the task has been satisfied or completed." While this definition captures the essence of the mentioned anecdotes, there are a number of examples that are similar to the description provided by Byrne and Bovair, but yet differ in a subtle way. Consider the following two examples:

- Setting a VCR with the appropriate times and channel but forgetting to press the "Record" button to set it to record.

- Setting an alarm clock with the desired time but forgetting to switch on the alarm.

Suppose the last step in each of the above example is carried out correctly (VCR set to "Record" and alarm switched on) the main goal is still not completed until the desired program is recorded or one is woken up at a particular time. However, they both involve omitting, arguably, a functionally isolated step at the end of an interaction. Using Curzon and Blandford's concept of interaction invariant, the VCR and alarm clock examples can be viewed as a second class of PCEs that omit the opening perturbation, but finish with a confirmatory step.

Defining a particular kind of error and classifying examples according to a precise error definition is usually open to slightly different interpretation. This is also part of the reason for the proposal of different terminologies such as "termination error" (Thimbleby, 1990) and "omitted secondary subgoal" (Young, 1994) describing the same error phenomenon as PCE describes. Nonetheless, the essence of the error in the number of examples mentioned above is that the final step of the task is secondary or peripheral to the main task goal, and needs to be carried out before moving on to the next task goal. The term PCE is used to refer to the error phenomenon in this thesis because the current work is largely motivated by the study carried out by Byrne and Bovair (1997).

In the above anecdotes a PCE might cause mere annoyance or inconvenience to our daily routine, however, in high-reliability and safety-critical systems, the consequences of human error can be severe. Rasmussen (1980) analysed 200 error incidents in nuclear power plants and identified that omission of functionally isolated steps, such as forgetting to set valves in the appropriate position, account for 34% of all incidents.

From a psychological point of view, it is important to understand the causes of error behaviour in terms of the underlying psychological mechanisms. Experimental studies of human error in the domain of cognitive psychology have investigated, for example, errors in statistical problem solving (Allwood, 1984), rule-based errors in arithmetic thinking (e.g. Payne & Squibb, 1990; Ben-Zeev, 1995) and slips in speech (Baars, 1992). However, there has been little experimental work investigating human error in the context of Human Computer Interaction (HCI), particularly on capturing error behaviours under controlled laboratory settings.

The scarce number of experimental studies of human error in HCI, as Senders and Moray (1991, p.2) suggest, is because "error is frequently considered only as a result or measure of some other variable, and not a phenomenon in its own right." One of the challenges in human error research is to design a suitable methodological paradigm that allows the desired error behaviour to be generated to a level that is adequate for comparing various independent manipulations statistically. However, recent research has attracted attention to the empirical studies of human error and suggests two approaches to investigate human error experimentally. One approach is to collect a large database of errorful and error-free performance in interactive behaviour and then subject the obtained data to fine-grain analysis (e.g. Gray, 2000). The other approach is to design an artificial task environment to provoke a particular kind of error behaviour in a controlled experimental setting (e.g. Byrne & Bovair, 1997).

The current thesis adopts the approach of having a controlled experimental setting to study various factors that provoke or mitigate the occurrence of PCE. While naturalistic approaches to studying human error have ecological validity, it is very difficult to isolate the identified variable(s) and study because there are often many things going on at the same time when carrying out a study "in the wild".

The first goal of the current thesis is to study PCE empirically in order to identify factors that provoke or mitigate the error. This goal is tackled by two series of empirical investigations: traditionally, research on PCE has been confined within the boundary of procedural tasks (e.g. Byrne & Bovair, 1997), however, informal observations from a number of anecdotes suggest that PCE may also occur in problem-solving situations (e.g. Polson, Lewis, Rieman, & Wharton, 1992). The first series of experiments (Chapter 3) is an investigation of PCE in problem-solving and the effects of a static visual cue on PCE.

Although interruption research generally finds that interruptions have disruptive effect on primary performance (e.g. Gillie & Broadbent, 1989; Hodgetts & Jones, 2005, 2006a and 2006b), there is no existing study investigating the effect of interruption position and duration on PCE. The second series of experiments (Chapter 6) examine the effects of interruption on PCE.

The second goal of the current thesis is to examine critically the existing theoretical accounts of PCE (Chapter 4) in order to gain theoretical insights into the error phenomenon. This is carried out through a meta-theoretical analysis bringing the seemingly disparate and incommensurable theoretical accounts under a coherent structure for comparative analysis. Although the analysis is not able to specify a complete model of PCE, it suggests that any adequate theoretical accounts should address three criteria: (1) how the task in question is carried out, (2) how PCE arise from the task performance behaviour, and (3) how the infrequent but persistent nature of the error is manifested. This meta-theoretical analysis also suggests Altmann and Trafton's (2002) activation-based goal memory model as a suitable theoretical basis to pursue the effect of interruption on PCE in a procedural task paradigm (Chapter 6).

The scope of the current thesis is described at the end of the next chapter, Chapter 2, which is a literature review of some of the major psychological research in human error in general, the empirical studies and theoretical accounts of PCE. Chapter 3 is the series of three experiments investigating PCE in problem-solving and the effect of static visual cue on PCE. Chapter 4 is the meta-theoretical analysis of extant PCE accounts. Chapter 5 is a literature review of existing research on interruption relevant to the current thesis. Chapter 6 is the series of four experiments examining the effects of interruption on PCE, and Chapter 7 is a concluding chapter summarising and discussing the implication of findings and contribution to knowledge of the current thesis.

Chapter 2

Literature review of human error research

1 Introduction

Major accidents involving safety-critical systems such as nuclear power plants (e.g. Three Mile Island incident in 1979), the space shuttle (e.g. Challenger disaster in 1986) and mass transit transport (e.g. Herald of Free Enterprise in 1987) are typically judged to be related to human error (Reason, 1990). However, human error as a research topic has not had as long a history as it deserves. The study of human error did not receive serious attention in the context of human-computer interaction (HCI), human factors (HF) and psychology until the early 80's. Moreover, traditionally, psychological research was primarily focused on error-free behaviour and not erroneous performance.

The first collective research effort on human error emerged in a conference held in Bellagio, Italy, sponsored by NATO in 1983. The conference involved participants from different academic disciplines: philosophy, psychology, mathematics and engineering. Discussions from the conference resulted in Senders and Moray's (1991) book — *Human error: Cause, prediction and reduction.* One of the main outcomes of the conference highlights an inherent difficulty in the study of human error, which is to arrive at an agreeable definition of the term "human error" across the different disciplines. However, Senders and Moray (1991) noted a generally accepted definition of human error among the different researchers and it usually contains the following attributes: 1) not intended by the actor; 2) not desired by a set of rules or an external observer; or 3) that led the task or system outside its acceptable limits (p. 25). These attributes include notions, such as, intention of an individual, the desirability of some performance outcome, and that the outcome is assessed against a task or system criterion. These definitional attributes are useful as a starting point in communicating what the term "human error" means in general.

The difficulty in arriving at a precise and universal definition of human error across different disciplines reflects the interdisciplinary approach to the research topic as a whole. The following sections serve as a brief review of the different perspectives or approaches in human error research, and the position adopted by the current thesis is then grounded in context.

2 Different perspectives in human error research

The need for a multi-disciplinary approach to research of human error is not only because it draws on a wide body of knowledge, but also because when searching for a causal explanation of human error there is no objective rule indicating when is the "best" point to stop the search (Rasmussen, Pejtersen & Goodstein, 1994). As Rasmussen pointed out, it is always possible to trace back one step further along a causal chain of events that is related to an error, and it is entirely dependent on an analyst's knowledge or approach to the investigation to decide the "cause" of the error after the fact. In general, there are two major perspectives in human error research and they can be described as the broad perspective and the narrow perspective. The different perspectives adopted not only affect how a researcher approaches a problem but also how the researcher explains an observed phenomenon. The following sections describe the two major perspectives in human error research and show which perspective the current thesis adopts.

2.1 A broad perspective

The nature of the broad perspective in human error research can be reflected by the definition of human error adopted by HF researchers such as Sanders and McCormick (1993): "Human error is an inappropriate or undesirable human decision or behaviour that reduces, or has the potential for reducing, effectiveness, safety, or system performance." (p. 656). This broad perspective emphasises that investigations of human error should take into consideration the entire system and not only the operator who is likely to be perceived as directly related to the error under investigation. The term "system" in this context usually refers to not only the physical technology but also a wider range of components in a work system, such as equipment designers, maintenance staff, supervisors, and managers. This broad perspective of human error is usually adopted by accident/incident investigators, whose primary interest is to attribute the cause of an error through retrospective analysis, or by system designers whose interest is to identify potential changes to a work system that will improve overall performance.

The broad perspective of human error is directly compatible with Reason's (1990) proposed distinction of latent error versus active error. Latent errors are similar to pathogens in the human body that may remain dormant for a long time before having any adverse effects on a system. Latent errors might be triggered or "activated" by a sequence of events or a combination of conditions, then weaken or break through a system's defence line and lead to failures. This type of error is associated more with people in the "background" of a work system such as decision-makers, system designers, managers and so on. For example, the nuclear plant disaster at Chernobyl in 1986 was not only because of the technology itself but more to do with a poor management structure, attitude to safety and poor training given to the operators (Reason, 1990).

2.2 A narrow perspective

Reason's notion of active error is very much representative of the narrow perspective of human error research. Active errors often result in immediate failure of a system and are usually concerned with immediate operators such as air traffic controllers, pilots, train drivers, nuclear plant operators etc. This is very similar to Hollnagel's (1998) notion of "erroneous actions", which also suggests that considerations apply only to the people who have direct involvement with the system interface. This narrow perspective of human error focuses its research agenda on the individual operator.

In general, the narrow perspective is mostly adopted by two branches of human error researchers defined by their different interests: evaluating human performance or understanding human performance. Human reliability engineers are primarily interested in human performance evaluation with the ultimate aim to develop predictive risk analysis of complex industrial installations. The risk analysis involved usually adopts some formal probabilistic frameworks. This approach is particularly important when human reliability is a crucial parameter, such as a monitoring role in a nuclear power plant since the operator is directly in a feedback control path in terms of the overall operation.

The other branch of human error research is characterised by psychologists who are interested in understanding human performance. This psychological approach views human error as "... an act which is counter-productive with respect to the person's private or subjective intentions of goals." (Rasmussen, Duncan & Leplat, 1987, p. 2). Moreover, human error is viewed as part of human behaviour and regarded as an essential part of skill development through trial-and-error in the environment (Rasmussen, 1982). In contrast to the reliability engineering approach, the psychological approach to human error not only focuses on observable behaviour but also on the underlying psychological mechanism(s) that give rise to the error behaviour.

In the context of HCI, tracing the cause of human error from a broad perspective can extend to the managerial level within an organisation, or even to the designer of the equipment or machine in question. However, the current thesis is concerned with understanding PCE from a cognitive perspective and a broad perspective investigation is not a suitable approach. A narrow perspective that focuses on investigating error behaviour arising from an individual (the operator) directly involved with a system interface is likely to be a more suitable approach in understanding PCE in terms of the cognitive causes. The current investigation adopts a psychological approach using laboratory-based experimentations, with the primary interest in identifying factors that provoke or mitigate PCE. The following sections review some of the major theoretical progress in human error research in terms of general psychological theories.

3 Classification of human error in various taxonomies

This section presents several dominant psychological theories of human error in general. The aim of this section is to exemplify the theoretical approaches adopted to understand human error behaviour, and also to set the error phenomenon under investigation, namely PCE, in the context of traditional human error classification.

Early theoretical effort in human error research has concentrated on devising classification systems to organise different errors that are observed or discussed into different categories. However, there is no single agreed taxonomy in human error research. Senders and Moray (1991) note that there are several ways in which an error can be classified; as a consequence, there are different levels of classification depending on which level the taxonomy adopts.

Senders and Moray (1991) summarise that there are three main levels of taxonomy: first, errors are classified in terms of their superficial manifestation. This is the simplest form of classification and is referred to as phenomenological taxonomies. Secondly, a taxonomy can have a lower level classification system in terms of the cognitive mechanism(s) involved. These taxonomies are usually grounded in psychological theories classifying errors according to mechanisms in human information processing. Thirdly, there are taxonomies that group errors according to biases or deep-rooted tendencies in decision-making process. For example, one typical psychological phenomenon in decision making is confirmation bias, which is a tendency to search for information that confirms ones hypothesis rather than disconfirming it (Wason & Johnson-Laird, 1972).

3.1 Phenomenological taxonomy

An example of a phenomenological taxonomy is a methodology proposed by Rouse and Rouse (1983) as a design and evaluation tool in relation to human performance in large-scale industrial installations such as aircrafts and power plants. Rouse and Rouse's taxonomy classifies operator errors assuming the information processing of the operator follows cognitive stages such as: observation of system state, choice of hypothesis, testing of hypothesis, choice of goal, choice of procedure and execution of procedure. It is assumed that during normal operations the human operator is engaged in observing the system state, and choosing and executing procedures. However, when a system is operating outside the pre-specified normal state, the human operator is then engaged in problem-solving involving choosing and testing hypotheses. Errors that may occur are classified under these various cognitive phases under normal or abnormal system operation. Such a classification scheme may be used as a practical tool to account for failures that might occur in industrial settings. However, the taxonomy does not presuppose a particular psychological theory of human performance; as a consequence, it does not help to understand human errors in terms of their psychological causes.

3.2 Skill-based, rule-based and knowledge-based human performance taxonomy

Rasmussen (1982) offers a framework for event analysis in complex industrial installations. Event analysis is a decomposition of the event into the features characterised by the taxonomy in order to understand the mismatch between human and system performance. Rasmussen's framework is a multifaceted description system, which describes human error in terms of cognitive domains as well as factors outside the human cognitive system such as situation-specific factors. Important to the current context is the cognitive components present in a framework of human performance. Rasmussen made a three level distinction in human performance: skill-based, rule-based and knowledge-based as the three categories one could use to classify human error.

Performance at the skill-based level involves dealing with highly familiar and routine tasks. Behaviour at this level is controlled by stored patterns or routines in the cognitive system that usually do not require direct conscious access by the skilled operator when performing the task. Performance at the skill-based level is primarily concerned with sensori-motor control and errors that occur at this level are very often attributed to faulty control of actions.

At the rule-based level, performance is concerned with familiar situations. The coordination of subroutines is carried out by stored rules in the cognitive system. In contrast to the automatic fashion of performance at the skill-based level, access to stored rules to find a solution often requires conscious mental effort at the rule-based level. Rule-based errors involve misrecognising the salient features of a situation and applying a wrong set of rules or misremembering a set of procedures.

In unfamiliar situations, one's behaviour is governed by the use of resource-intensive analytical processes and stored knowledge. Performance in unfamiliar situations is at the knowledge-based level in which the operator is problem solving. Errors that happen at the knowledge-based level are usually the result of incomplete or incorrect knowledge. Rasmussen pointed out that performance at the knowledge-based level depends on details of the situation and the task context; therefore, error mechanisms at this level can only be inferred from detailed studies with, for example, verbal protocol analysis.

The three levels of performance together highlight the notion of familiarity of the operator with the environment and the task. Shortcuts might happen between certain performance levels and result in, for example, situation-specific stereotype fixation in which the confronted situation bypasses the labour-intensive knowledge-level processing and leads directly to the selection of 'appropriate' procedures.

In terms of describing PCE within Rasmussen's framework of human performance, the error is best described at the rule-based level performance. The occurrence of the error is not due to incomplete or incorrect knowledge of the task, and is not because of mismatch of sensori-motor activity between the human and the artefact involved. Rasmussen (1987) commented that a common category of error at the rule-based level is the omission of an isolated act, which is not a logical part of the main task sequence. Such omission is attributed to lapses of memory because the nature of the isolated act is not an integral part of a larger memory structure.

3.3 Distinction between slips and mistakes

Reason (1990) approximates Rasmussen's skill-rule-knowledge classification of human performance distinguishing between three error types: skill-based (SB) slips, rule-based (RB) mistakes and knowledge-based (KB) mistakes. Reason's three-level framework of performance is slightly different to Rasmussen's in that, like KB-level processing, processing at RB-level is also viewed as involved in problem-solving activity. Moreover, SB-level processing is usually involved in routine and non-problematic activities. Reason defined slips "*as errors which result from some failure in the execution and/or storage stage of an action sequence, regardless of whether or not the plan which guided them was adequate to achieve its objective.*" (p. 9). Mistakes are defined as "*deficiencies or failures in the judgemental and/or inferential processes involved in the selection of an objective or in the specification of the means to achieve it, irrespective of whether or not the actions directed by this decision-scheme run according to plan.*" (p. 9).

Reason explains how various *types* of errors may arise through transitions at the three different levels of performance. Processing is always sequential and starts at the SB level which is responsible for routine activities. At this point, attentional monitoring is the main mechanism for keeping track of one's progress and, if the activity starts deviating from routine actions then RB-level processing will be called into play. The cognitive system will then search through the pre-established knowledge structures, schemata, for the closest matching rule in relation to the environment. If no approximate rules can be found then the effortful, computationally demanding processing at the KB level will be required. It is suggested that SB slips are distinguished between those caused by inattention and those by overattention. The application of bad rules or misapplication of good rules can lead to the occurrence of RB mistakes. Finally, KB mistakes usually arise because of 'bounded rationality' or inaccurate/incomplete representation of the problem space, which could lead to different types of error such as confirmation bias, over-confidence and working memory limitations.

The various types of errors have also been proposed, by Reason, as results of two basic error forms: schema-matching and frequency-bias. Reason suggests that the two error forms are a basis for human errors, which arise from the same cognitive processes that are responsible for normal learning and performance of cognitive skills.

In light of Reason's classification scheme, PCE can be viewed as a slip because it is not due to deficiency in knowledge or planning of an action sequence but more to do with a failure in the execution or storage of the action sequence. Although Reason categorised slips at the SB level whereas Rasmussen would categorise them at the RB level, these views are not conflicting because both views are describing the same notion, namely, human performance in routine, familiar situations. However, the use of different terminologies reflects something more fundamental in the ambiguity of classifying errors. As Gray (2004) comments "the knowledge-based, rule-based, and slip-based approach to errors is neither as neat and clean nor as theory-based as it may first appear. Whether an error is classified as skill-based, rule-based, or knowledge-based may depend more on the level of analysis than on its ontogeny." (p. 3).

Norman is another proponent who makes a distinction between slips and mistakes. The distinction is succinctly summarised as "An error in the intention is called a mistake. An error in carrying out the intention is called a slip." (Norman, 1983, p. 254). Norman (1981) proposed a model, the activation-trigger-schema system (ATS), to classify a variety of slips into different categories. The general principles of ATS state that there are schemas, which are memory units or an organised body of knowledge, and they are organised in a hierarchical structure. Schemas at the highest level are parent schema and schemas at lower levels are child schemas. There are activation levels associated with the schemas and interactions between schemas are possible through the activation mechanism. When the parent schemas are activated, this represents the formation of an intention, which is carried out by child schemas. Sources of activation can come from multiple sources such as the external world, internal processing (e.g. other associated memory structures) or familiar habits. More than one schema can be active at any one time and the selection of a schema is governed by a triggering condition. The triggering condition does not need to be an exact match for schema selection. Therefore, the selection of a schema is a trade-off between activation level and goodness-of-match of the triggering condition.

Norman suggests that ATS specifies three main sources contributing to action slips: fault in the formation of intention, fault in activation and fault in triggering. Faults occurring in the formation of intention can result in mode errors or description errors. Mode errors involve incorrectly classifying a situation leading to performing actions that are inappropriate for the actual situation. Description errors usually involve insufficient specification of the intention resulting in behaviour inappropriate for the current situation.

Faulty activation can have two forms of manifestations: unintentional activation or loss of activation. Unintentional activation of schemas not related to a desired action sequence can result in errors such as capture errors. Capture errors involve the control of an action sequence captured by a more practised, familiar sequence. A schema may lose activation due to decay or interference and the loss of activation can result in errors such as skipping a required step in an action sequence.

Slips occurring due to faulty triggering can also have two forms of manifestation: false triggering and failure to trigger. False triggering involves selecting an active schema at an inappropriate time leading to errors such as blends: performance of a combination of two action sequences. Failure to trigger is when an active schema does not get invoked and this can be because the schema has a very low initial activation level or the triggering condition is poorly specified.

By viewing PCE in light of Norman's taxonomy, it is reasonable to describe the error as a slip in terms of omitting a step in an action sequence. However, the occurrence of PCE can be due to faulty activation resulting in a loss of activation or faulty triggering resulting in a failure to trigger the appropriate action, and it is not clear which is the best way to classify accordingly. This is not a problem specific to PCE because classifying human errors in general is not as clear-cut as one might think, as Norman (1981, p.7) commented "As is usual, these errors [referring to mode errors and description errors] most likely have several causes." Therefore, there are often some ambiguities in classifying some errors into different categories. Nonetheless, the manifestation of PCE is more likely to resemble the characteristics of a slip that it is not because of insufficient or incorrect knowledge of a task but a momentary lapse in memory during the execution of the action sequence.

While classification using taxonomies is a good starting point, if not a necessary one, in organising a huge variety of data about a phenomenon, human error taxonomies suffer from several problems: first, as mentioned earlier, classification of an error is often not as clear-cut as one would hope for and there might be several categories that an error might fit. Secondly, related to the ambiguous nature of error classification, Byrne (2003) points out that in terms of explanations based on causal mechanisms, taxonomic frameworks have the problem of heterogeneity; that is, grouping unlike things together and like things apart. Two erroneous behaviours might be grouped under a same kind of error, but the underlying causal mechanisms responsible for the two surface behaviours might have completely different nature and vice versa. Thirdly, and most importantly, taxonomies are only useful for analysis of errors after the event; they cannot be used to predict the occurrence of errors. As Gray (2004, p. 3) points out, "Although some of these taxonomies rely on cognitive theory as the basis of their classifications, all lack the mechanisms to predict errors." This is likely because the theories underlying the taxonomies are only specified in terms of some general principles of operation. As Norman (1981, p.4) points out "More detailed specification is, of course, required for the understanding of any specific action sequence, but at the moment there are not sufficient data to justify more details."

Placing PCE in the context of taxonomies, which are tied with general theories of human error, helps distinguish the error from knowledge-based mistakes. The occurrence of PCE resembles more closely the characteristics of a slip: involving failure in the execution of an action sequence rather than the incorrect or incomplete specification of knowledge of the task. While general theories of human error fail to generate testable predictions of PCE, there have been some theoretical efforts specific to PCE and the following section introduces those existing theoretical approaches to the error phenomenon. This is then followed by a review of the empirical studies related to PCE.

4 Extant theoretical approaches to PCE

There are several theoretical approaches to account for the phenomenon of PCE: the supergoal kill-off approach, the Soar approach, the activation-based goal memory approach, and the working memory capacity approach. This section serves as an introduction to each of the PCE accounts, a more detailed analysis of the accounts is presented in Chapter 4.

4.1 Supergoal kill-off approach

Polson, Lewis, Rieman and Wharton (1992) developed a model of learning by exploration within the construction-integration (C-I) model (Kintsch, 1988). Polson et al. applied their model to explain the phenomenon of PCE, but instead of calling it PCE, they termed it supergoal kill-off.

The supergoal kill-off approach assumes that goals are organised in a hierarchical structure with the top-level goal representing the overall task goal, which can be decomposed into subgoals, and these subgoals can be further decomposed into the lowest-level goals representing executable actions. The model specifies that the hierarchy of goal is dynamic and when an action is executed the goal structure is revised. This highlights the important role of feedback from the environment as it helps the model to generate new goals when the goal structure is revised.

Goals and subgoals are linked together by activation links and activation flows through these links from the top-level goal to its subgoals and down to the corresponding action nodes. All subgoals receive activation from the top-level goal as long as it remains active. When a subgoal receives sufficient activation its corresponding action will then be executed. Each goal or subgoal has an associated "done-it" node attached to it by an inhibitory link. The "done-it" node gets activated once its associated goal is accomplished and, through the inhibitory link, the goal is deactivated and activation flows to and from this goal are cut off. The flow of activation is governed by the presence of "and-then" nodes, which act as activation gates between subgoals. An "and-then" node receives activation from the "done-it" node of a subgoal and spreads activation to the next subgoal. This "and-then" control pattern means subgoals are executed in an ordered sequence such that the second subgoal can only be carried out when the first subgoal is accomplished and the top-level goal is active. Polson et al. suggest that this type of goal execution in an ordered sequence is typical in procedural tasks such as operating an ATM.

The supergoal kill-off model specifies that when a subgoal is accomplished in the "and-then" control structure, the subgoal and the top-level goal are checked for their similarity. Similar goals can have their respective "done-it" nodes associatively connected, so the fulfilment of a goal can cause a similar goal to deactivate. The supergoal kill-off approach explains PCE in terms of similarity between the top-level goal and its subgoal, which causes premature deactivation of the top-level goal. In the example of the ATM case, the subgoal of collecting cash might be very similar to the top-level goal of withdrawing cash. When the subgoal is accomplished, this causes deactivation of the top-level goal and activation supply to the PC subgoal of collecting the card is terminated and its action not carried out.

However, the explanation offered by the supergoal kill-off approach is problematic in two ways: first, the approach does not specify how the notion of similarity should be operationlised, that is, it is difficult to determine what it means for two goals to be similar. Secondly, even if the level of similarity can be determined, the approach would then predict PCE to occur all the time when similarity occurs and no error at all when dissimilarity occurs.

4.2 Soar approach

Young (1994) proposed an account of PCE within the Soar cognitive architecture and suggested that the error phenomenon can be explained by the inherent property of how subgoals get initiated and terminated within the architecture.

In Soar, behaviour of a model emerges by applying operators to states in problem spaces, and the choice and application of operators are determined by knowledge in a long-term memory repository. When no more progress towards a goal can be made within a problem space, an impasse occurs and Soar then initiates a subgoal to acquire knowledge in other problem spaces. When the impasse is resolved, control is then resumed to the main goal but the subgoal and its associated structure then disappear. Young suggested that there is asymmetry between initiation and termination of a subgoal within Soar, which is important in explaining how a post-completion (PC) subgoal is omitted. The occurrence of an impasse allows Soar to initiate a subgoal to trigger activity, however, the termination of an impasse is a non-event such that a terminated subgoal simply disappears and cannot be used to trigger further activity.

In the case of withdrawing cash from an ATM, the Soar approach explains forgetting to take the card back at the end by applying the asymmetry property. The Soar approach suggests that an impasse occurs when using the ATM and a subgoal is set up to work through the necessary procedures to withdraw cash. The impasse is then resolved when the cash is withdrawn and the associated subgoal of collecting the card simply disappears and is not executed.

The Soar approach is different to the supergoal kill-off approach in that it does not rely on theoretical constructs such as activation but an inherent property within the architecture. However, the asymmetry is only a starting point of the explanation to PCE because it does not account for how the task is ever performed without committing the error. Young further suggested that remembering of the PC subgoal might be achieved by learning appropriate cues, such as self-reminding or reliance of external cues. Moreover, the usage of these cues or strategies are necessarily fragile such that they could be disrupted easily by, for example, distracting conditions. Although details of the compensatory strategies to overcome the error and disruptive conditions are not specified further, this multi-level structure of explanation is an important insight which can be used to address the problem of having a deterministic model generating error behaviour either all the time or no error at all.

4.3 Activation-based goal memory approach

The activation-based goal memory (AGM) model (Altmann & Trafton, 1999; 2002) was developed to address some of the fundamental limitations of traditional approaches to goal memory. The traditional view of goal memory in cognitive architectures assumes a stack-like structure (Anderson & Lebiere, 1998; Newell, 1990). A goal-stack memory exhibits error-free behaviour in goal selection such that a goal order is perfectly preserved. In contrast, the AGM model treats goals as ordinary memory elements in that they are subject to various constraints such as decay, noisy retrieval that is vulnerable to interference, and cognitive costs in active maintenance.

The AGM model is implemented in the ACT-R cognitive architecture (Anderson & Lebiere, 1998), with its goal stack "switched off", to model the goal selection process of the Tower of Hanoi task, which is a task often used in problems-solving research (e.g. Simon, 1975). There are three main components in the AGM model: the interference level, the strengthening process and the priming process. The interference level is a gateway determining the most active goal in memory. Goals in the AGM model undergo a gradual decay process that is time-based. This means that decaying goals can act as distractors to the retrieval of a target goal forming a mental clutter. The interference level is effectively the mean activation level of the most active distractor. In order for a goal to direct behaviour, its activation level must be above the interference level.

The strengthening process specifies how a new goal gains activation to overcome interference imposed by old goals. During the process of strengthening, a goal is effectively being encoded in memory, in other words, attention is being focused on it. The strengthening process determines how long should be spent in encoding a goal because spending too little time means the goal does not get enough activation, however, strengthening for too long increases the interference level and affects retrieval of subsequent goals. Once a goal is strengthened above the interference level and starts directing behaviour, it begins to decay.

An old goal has to overcome retroactive interference in order to be retrieved successfully and this is specified by the priming process. An old goal can receive a boost in its activation level from its context, and this context manifests as external cues in the environment or internal mental cues. For a cue to be effective, association between the cue and the target goal must be learned in the first place. The priming process highlights that the process of goal selection, which can be driven by environmental cues or mental cues. The AGM model suggests that procedural skills can be viewed as a chain of associative steps; the execution of each step associatively primes the next step.

Altmann and Trafton describe the occurrence of PCE within the AGM model. The AGM model suggests that when a subgoal is suspended it undergoes decay, and this implies that PCE should occur by default. This is because the PC subgoal is the last subgoal in a task and should have decayed the most. However, Altmann and Trafton suggested that the fact PCE is often avoided is because of the priming process; it is postulated that associative cueing occurs between the PC subgoal and its preceding subgoal. The execution of the penultimate subgoal serves as a cue to prime the PC subgoal in procedural tasks. However, this PCE approach is problematic in its assumption that the PC subgoal should have decayed the most. By the same logic, given a long enough task, subgoals that are in proximity to the PC subgoals should also have decayed and be omitted. Moreover, the account ultimately provides a description of how PCE is avoided rather than how the error is made.

4.4 Working memory capacity approach

Byrne and Bovair (1997) proposed a working memory capacity account of PCE predicting that PCE is more likely to occur under high working memory demand. The account is developed in the CAPS cognitive architecture (Just & Carpenter, 1992), which has an explicit mechanism of working memory capacity.

In the CAPS model of PCE, active goals are held in working memory and in order to remain active a goal has to have an activation value above a certain threshold. Like the supergoal kill-off approach in C-I, activation is propagated from the top-level goal to its subgoal and the activation propagation process is in place as long as the top-level goal remains active in working memory. When the cognitive system is operating above its working memory capacity, activation levels for both maintenance and propagation get scaled back.

Satisfied goals decay in the CAPS model, and the decay process is dependent on working memory load rather than time as in the AGM model. Satisfied goals only get displaced from working memory when there are other items to absorb their activation. In explaining the occurrence of PCE, the CAPS model postulates the following: when the cognitive system is operating within its working memory capacity, the main goal remains active in working memory even after being accomplished and there is no cut-back in maintaining and propagating activation to the PC subgoal. Therefore, the PC subgoal becomes active and gets executed. However, when the working memory load exceeds the cognitive system's capacity, the satisfied main goal gets displaced from working memory and also the amount of activation propagated to the PC subgoal is cut back. As a consequence, the PC subgoal fails to receive enough boosts in activation before it is displaced from working memory.

The working memory capacity account is problematic in two ways: first, as Byrne and Bovair pointed out, since satisfied goals in the CAPS model can still be carried out it is necessary to impose a minimum amount of working memory load in order to get the satisfied subgoals displaced from working memory. This is to ensure subgoals are carried out in the correct sequence without repeated execution of a previously accomplished goal. However, correct behaviour usually occurs without the constraint of a fixed minimal working memory load. Secondly, although the working memory capacity specifies the condition under which a PCE is more likely to occur, the model generates deterministic error behaviour, that is it makes the error all the time under high working memory load and no errors at all when working memory load is low. This will be discussed in more detail in Chapter 4.

Apart from the working memory capacity approach, the existing theoretical approaches to PCE are all verbal descriptions of the error phenomenon. The explanation offered by Byrne and Bovair is the most comprehensive in that a computational model is developed to capture one of the main empirically tested aspects of PCE, namely, working memory load. The theoretical progress of PCE is still in its infancy and this is mainly because there are only a limited number of empirical studies looking specifically at PCE. This is also partially related to the lack of experimental investigation of human error in general. The following sections provide an overview of the main empirical approaches in studying human error, and review the experimental studies of PCE.

5 Empirical studies of human error

The topic of human error is broad and spans many different disciplines. However, the study of human error is shallow and the main reason, pointed out by Gray (2004, p. 1), is that "... although many researchers collect error data, there is no established research tradition to experimentally manipulate and study error as a phenomenon." One of the difficulties in studying errors systematically in experimental settings is because error behaviour is rare, as Gray (2004, p. 4) further comments that "This rarity may have encouraged the naturalistic approach in which researchers and their confederates carry around notebooks with the goal of noting and recording the occasional error." While naturalistic approaches are useful in documenting the frequency of error occurrences, reports of errors after the fact usually do not capture the essential context in which the error occurred. Gray (2004, p. 4) contends that "... they [naturalistic approaches] have not been particularly productive in understanding the cognitive mechanisms that determine the nature, detection, and correction of errors in interactive behaviour."

5.1 Two empirical approaches to human error

The experimental study of human error can be divided into two main approaches: the first approach involves collecting a large database of error behaviour as well as error-free behaviour and subject them to fine-detailed analysis. This approach has been adopted in investigation of errors in arithmetic thinking (e.g. Payne & Squibb, 1990; Ben-Zeev, 1995) as well as speech errors (e.g. Baars, 1992). A recent study carried out by Gray (2000) has adopted this fine-detailed approach to examine various errors in the interactive behaviour of VCR programming. In Gray's study, the measurement unit of interactive behaviour is number of keypresses made on a simulated VCR. Although only 4% of the total keypresses collected from the participants were identified as errors, had these incorrect behaviours not been corrected they would have led to a failure in recording the desired TV programs successfully. Gray was able to describe and explain the obtained errors under a cognitive framework, which will be described in more detail later in the chapter.

The second empirical approach of studying human error is to elicit a particular error in a controlled laboratory environment. This approach is difficult for two reasons: first, participants arriving at an experimental session are likely to be more conscious about their performance than usual, and self-monitoring mechanisms can be the main obstacle to obtaining the desired error behaviour. Secondly, as Sellen and Norman (1992, p. 335), comment, "Another difficulty inherent in laboratory-induced errors is the low frequency of errors in artificial tasks. It is often impossible to obtain a large enough sample of errors in a reasonable time to conduct statistical tests. This is especially true when one is looking for a particular kind of error."

Despite the difficulties of this experimental approach, there are a number of studies that successfully elicit particular kinds of error in laboratory settings. For example, capture error in the visual domain has been successfully generated and examined under controlled conditions (Larson & Perry, 1999). Visual capture error is the erroneous behaviour driven by external context rather than by one's internally generated goal. Larson and Perry studied susceptibility to visual capture in relation to individual differences in everyday error proneness and working memory capacity using the antisaccade paradigm, which measures ones ability to inhibit reflexive responses to visual stimuli in terms of eye movements. It was found that error-prone participants were more susceptible to visual attention capture than less error-prone participants, suggesting that error-prone individuals find it harder to suppress distraction from external stimuli that dominate internal goals. Moreover, working memory capacity was found to be uncorrelated with visual capture susceptibility. However, Larson and Perry pointed out that the sample in their study had generally high working memory scores and this might have contributed to the non-significant finding.

Instances of mode error have also been studied in controlled experimental conditions. Monk (1986) analysed mode error from a user-centred perspective and examined the effect of keying-contingent sound on the error occurrences. A computer-based game paradigm, with two possible modes of operation, was designed to elicit mode errors. Two conditions of sound feedback upon key-presses were tested: a control condition that consisted of a single pitch level tone across the two operation modes in the computer game; and an experimental condition, which had two different pitch tones (high and medium) for the two different modes of operation. Results from the study suggested that despite the low overall frequency of mode error generated, participants in the experimental condition made about 1/3 of the mode errors obtained in the control condition. The effectiveness of differential keying-contingent sounds to mitigate mode errors was attributed to an increased general awareness of which mode one is in, and thereby reducing the cognitive burden in remembering which mode is currently active.

Another kind of error known as the unselected window (USW) error has also been successfully generated and studied using an artificial task (Lee, 1992). USW error occurs in multi-window environment in computer use and the error is an omission to select a desired window before interacting with it; so input gets sent into a "wrong" window instead. USW error is an action slip and is also classified as very similar to, or simply a specific case of, omitted secondary subgoal (Young, 1994). In Lee's study, two variants of USW errors were generated: one variant occurred within the same computer terminal during switching between different windows; the other variant occurred between different computer terminals when the participant completed an external task and returned to a multi-window environment at the initial terminal.

Lee designed an artificial task involving multi-window usage to elicit USW error and examined the effect of time course and dynamic visual feedback on the error. The computer-based task required participants to buy or sell shares according to some simple decision rules. The effect of visual feedback was tested by two conditions: dynamic condition with a fizzy border appearing around the should-be-selected window after 25 seconds of inactivity, and the static condition, which had a black border around the desired window. An interruption task was also implemented in a separate computer terminal requiring participants to copy down a 14-digit random number every 90 seconds. The participants in the study were required to perform trials with the interrupting task and without the interrupting task in order to simulate the two variants of USW. In order to assess the effect of time course, the participants were tested for three one-hour sessions spread over three days within a week. The results showed that there was initial decrease in USW error after the first day of testing; however, the error persisted throughout the second and third day though no differences in the error occurrence were found between these two days. There was also no reliable difference between the interrupted and non-interrupted condition in terms of USW error. Moreover, dynamic visual feedback was found to be effective in reducing the number of USW error whereas static feedback had no effect on error reduction.

Lee explained the effectiveness of dynamic visual feedback in terms of its attention-capturing property and that it increased participants' general awareness of window selection. Lee (1993) extended the study to examine the effect of a longer time course on USW error. The study involved the use of static visual feedback only and increased the time course to four one-hour testing sessions on different days spreading over a week. It was found that USW error decreased after the first day of testing and remained at the same low rate throughout the second and third day; however, the error occurrence increased significantly in the last day of testing. Lee accounted for the finding by postulating that there was considerable cognitive cost in remembering to select a window since static visual feedback did not help in mitigating USW error. As a result, participants rely on error correction rather than error prevention strategy to avoid the cognitive overhead as there were no dire consequences associated with error commission in the study.

Although there are a limited number of studies of laboratory-induced error in interactive behaviour, the studies mentioned above are the few examples that suggest it is possible to generate slip-like errors in a laboratory setting with the appropriate choices of task paradigm. Recently, PCE has also attracted attention, and has been artificially generated and studied empirically. The following section reviews those studies that are directly relevant to PCE.

6 Empirical studies of PCE

In studying the nature, detection and correction of error behaviour in a rule-based procedural task, namely, VCR programming, Gray (2000) adopted the approach of collecting a large database of error-free and error behaviour and analysed them in fine-grain detail. Using a simulated VCR, Gray collected data from 9 participants who were asked to program the VCR and their actions were recorded as action protocols. In the analysis, Gray adopted a cognitive framework guided by three least-effort cognitive principles to describe and explain the correct and incorrect interactive behaviour. The three cognitive principles were 1) least-effort in operating the device such as minimising cost in terms of perceptual-motor level; 2) least-effort in mapping existing prior knowledge to device knowledge such that participants would assume the VCR works in a way according to their existing understanding; and 3) a least-effort strategy of display-based difference-reduction, this is a strategy to reduce the distance between one's current state and the target goal state by offloading associated cognitive burden to the device's display. For example, in setting the VCR clock to a desired time the display-based difference-reduction strategy allows one to use the display information to hill-climb to the desired value without having to keep track of the number of key-presses one needed to carry out in order to reach the target.

The main idea behind the three least-effort principles is the notion that cognitive effort is minimised whenever possible. Gray then implemented these three cognitive principles in a computational model using the ACT-R cognitive architecture and used the model to trace through each individual participant's action protocol. Gray used the model-tracing technique to identify matches and mismatches between the model and the human behaviour, mismatches were identified as errors and subjected to further analysis to examine if they were "true" errors or limits of the model.

It was found that the model accounts for the vast majority of correct behaviour, and the display-based difference-reduction strategy accounts well for most of the detection and correction of error behaviour. Certain limitations of the model were also revealed; for example, the model had perfect memory of programme information, and was never distracted by other items on the VCR display. Gray was able to categorise the mismatches obtained from model-tracing into two kinds of error: push error and pop error. A push error is an error in adopting a goal or subgoal; for example, Gray accounts for display-induced mode errors suggesting that these errors arise because a certain goal needs to be done; however, information on the device's display is not sufficient to indicate one's place in the current mode. A pop error is an error in terminating a goal or subgoal and it can be further divided into premature pop, in which one stops before completing a goal, or postponed pop, in which one continues executing a goal even though the goal is completed.

Although Gray's study is not a specific investigation into PCE, the study documented the occurrences of the error phenomenon. To successfully program the simulated VCR, participants were required to set various parameters such as date, channel and time to the desired values and then the last step was to set the VCR to record. Gray (2000, p. 244) noted that "... device-specific goals are the hardest for participants to remember to accomplish. In this study, many of the trials that were not successful failed because of the post-completion error of forgetting to set the VCR to PROG REC." A total of 26 errors, which had prevented the VCR from successful recording, were identified and there were 6 PCEs (23%). Gray's analysis suggests PCE is a premature pop error: the early termination of a goal leads to omission of the PC subgoal. The display-based difference-reduction perspective suggests that the difference between the current goal state and target goal state is minimal when the various parameters (date, time, channel) on the VCR are set, and this lures one to terminate the task prematurely, as a result, forgetting to set the VCR to record. However, this explanation of PCE is problematic because it does not explain how the PC action is remembered and gets executed most of the time. However, Gray's investigation is not focused solely on PCE and he also admitted that the model does not account for factors such as working memory load, which is found to have an effect on PCE (Byrne & Bovair, 1997).

The first experimental paradigm, using the approach of generating a particular error under controlled conditions, to study PCE was carried out by Byrne and Bovair (1997). Byrne and Bovair carried out two experiments and examined, specifically, the effect of working memory load on PCE. In the first experiment, the main aim was to set up an artificial task that is sufficiently complex to generate systematic occurrences of PCE. A complex fictional computer-based game, called the Phaser task, was designed. The objective of the Phaser task is to destroy enemy ships and operation of the Phaser involves following a pre-defined set of procedures with various subgoals. To fire the Phaser, a component called "Tracker" needs to switched on and when the enemy ship is destroyed the "Tracker" needs to be switched off. Forgetting to switch off the "Tracker" is classified as a PCE. A similar task but simpler with fewer subgoals was also designed to compare task performance. Participants were trained and tested on both tasks. It was found that the simpler task was unable to produce many PCEs, however, the complex Phaser task was able to generate a high level of PCEs. A systematicity threshold of error occurrence was calculated as a ratio of obtained number of errors to the number of opportunities for that error to occur. The PCE rate obtained in the experiment had a 9.3% (13 out of 140 opportunities) systematicity score which is higher than the 5% threshold needed to classify an error occurrence as systematic. Byrne and Bovair suggest that the Phaser task is sufficiently complex to elicit PCE for further investigation.

In the second experiment, the effect of working memory manipulation was tested. In addition to the Phaser task, a similarly complex task, Transporter task, with the same number of subgoals and task structure was also included in the experiment to establish that the occurrence of PCE is not specific to the Phaser task. The Transporter task differs from the Phaser task in terms of visual appearance and has a shorter tracking subtask. In a condition where working memory is taxed, participants were required to carry out a concurrent letter recall task. The recall task involved continuous broadcasting of randomly generated letters every 3 seconds and participants were required to recall the last three heard letters at a random time between 9 to 45 seconds. Participants in the no working memory demand condition were not required to perform the letter recall task. Working memory capacity of each participant was also measured, from which participants were divided into two groups: low and high capacity. It was found that the overall systematicity of PCE reached a level of 48% (294 out of 610) across tasks. There was a general increase in overall errors with the Phaser and Transporter tasks when low-capacity participants performed the tasks with high working memory load. It was observed that PCE was rarely made by high-capacity participants with no working memory load but low-capacity participants with high working memory load produced reliably more PCE. Moreover, there were more PCEs made in the Transporter task than the Phaser task. Byrne and Bovair explained the difference in terms of the different tracking times in the two tasks; it was suggested that the relatively longer tracking time in the Phaser task might have acted as a buffer period freeing up working memory capacity, and, as a consequence, leading to fewer PCEs than the Transporter task, which has a shorter tracking time. Byrne and Bovair concluded and proposed a working memory capacity account to explain the effect, suggesting that under high working load the PC subgoal is more likely to be displaced from working memory and left forgotten when the main task goal is accomplished.

The effect of working memory load on PCE has also been replicated by a different study using a computer-based game paradigm (Mortenson, 2003). A robbery game was set in a context where participants were required to collect items from different houses; the items had to be remembered and recalled once a robbery was completed. A device was required to be switched on to search for the items and the PC step required the device to be switched off after the recall of the item collection. The effect of working memory load was manipulated by the different number of items to be collected and recalled; participants were required to collect 1, 3, 5 or 7 items. Consistent with results from Byrne and Bovair (1997), it was found that more PCEs were committed with higher working memory demand (5 or 7 items) than with lower working memory demand (1 or 3 items).

Factors related to mitigating PCE occurrences have also been investigated. Chung and Byrne (2004) examined the effect of visual cues on PCE, specifically, performance from a dynamic just-in-time cue (occurring just-before the PC step) and a static asynchronous cue (occurring earlier in the task) was compared. The same Phaser task (in Byrne & Bovair, 1997) was used alongside a simpler task with fewer subgoals. There were three visual cue conditions: first, a dynamic condition involving blinking arrows pointing to the PC step button and the arrows appearing just-before the PC step needed to be executed; secondly, a static condition where the PC step is highlighted and a message indicating the "Tracker" mode is on and this cue appeared when the "Tracker" is switched on; thirdly, a control condition with no visual cues regarding the PC step. Participants in the experiment were also required to perform the concurrent letter recall task to increase working memory load. It was found that the dynamic just-in-time cue was effective in completely eliminating PCE, however, the static asynchronous cue had no reliable effect on PCE compared to the control condition. The simpler task generated a low PCE rate overall but the dynamic cue was still able to eliminate all PCE. Chung and Byrne explained their finding in terms of the AGM model (Altmann & Trafton, 1999) suggesting that the just-in-time property of a visual cue is important in determining its effectiveness in mitigating PCE. An asynchronous cue occurring earlier in a task adds memory load to the participants while performing intervening subgoals, and the effect of the cue is likely to be masked by the other subgoals. It was also suggested that the dynamic nature of a visual cue helped in capturing participants' attention in acknowledging the presence of the cue.

The effects of motivational factors on PCE was also examined (Byrne & Davis, in press). Using the same Phaser task paradigm as Byrne and Bovair (1997), Byrne and Davis carried out two experiments to investigate the effect of practice and various motivational factors on PCE. In the first experiment, participants were required to go through three separate testing sessions with the Phaser task. In the experiment, a non-postcompletion (non-PC) version of the Phaser was also included, in which the PC step of switching off the "Tracker" had to be carried out in order to receive feedback about goal completion. All participants were also required to perform the concurrent letter recall task to impose a working memory load. The result suggested that significantly more PCEs were obtained in the PC version of the task than the non-PC version in the first testing session. Occurrence of PCE in the PC task declined significantly throughout the second and third testing, whereas the PCE rate in the non-PC task remained low throughout all testing sessions. Byrne and Bovair suggest that while more testing sessions might need to be included to test the effect of prolonged practice on PCE, this initial finding indicates practice can help mitigate the error.

In the second experiment, Byrne and Davis used the same experimental paradigm to examine the effect of various motivational strategies in mitigating the error. The motivational strategies include, praising good performance, reporting of poor performance and re-training. A task redesign condition was also tested in which the Phaser task was switched from the PC version to the non-PC version when the intervention was applied. Task performances in the various intervention conditions were compared to a baseline condition in which no intervention was applied. The various interventions were introduced in the middle of the second testing sessions. Task performance was compared between the first and the third testing sessions. It was found that occurrences of PCE decreased in all intervention conditions and across testing sessions, and that all interventions were equally effective in alleviating the error. Moreover, the rate of error reduction in the intervention conditions did not differ from the baseline condition. Byrne and Bovair concluded that motivational factors that were tested were not more effective in combating PCE than practice alone. Although the interventions did not differentiate statistically in error reduction, the redesign condition obtained a complete elimination of PCE. Results from this study suggest that PCE has a persistent nature in occurrence and the complete elimination of the error might be best achieved by avoiding a task structure with a PC subgoal.

Reason (2002) carried out a survey of errors made in photocopying and analysed the PC step (collecting the original) in terms of its task structure characteristics. In the survey, 95 undergraduates and academic staff were given questionnaires about various possible errors that could be made in photocopying. Each item on the questionnaire had a 7-point scale (0 = never, 6 = nearly all the time) and the participants were asked to rate each item accordingly. It was found that the PCE, leaving the last page of the original, had the highest score among all other omission errors (e.g. fail to copy all pages) with a mean of 2.18 and a standard deviation of 1.56. The result suggests that the PCE does not always occur but it has a persistent occurrence. Reason then analysed the photocopying task and identified a number of characteristics that make the PC step particularly vulnerable to omission:

1) The output of the last copy page acts as a *false completion signal* and can be a miscue in cueing one to terminate the task before all necessary steps are carried out.
2) The strength of the false completion signal is particularly powerful at the end of the task because ones attention starts to shift to a subsequent task leading to "premature exits".
3) The PC step is functionally isolated in relation to the entire task.
4) There is no cue visible in reminding one about collecting the original since the original is obscured from view by the photocopier's lid.

However, the survey data in Reason's study might suffer from biases, for instance people might tend to remember forgetting to collect the original more than other omission errors. A naturalistic approach is prone to biases imposed by factors such as unreliable memory recall of incidents. Alternatively, a longitudinal diary study might provide a more reliable source of data. However, the survey data suggests that forgetting to collect the original is infrequent but is persistent in occurrence. Furthermore, through the analytic approach in terms of task characteristics, Reason offered some useful insights into the elements that make the PC step vulnerable to omission.

7 Summary

In sum, from early taxonomic approaches to human error, PCE can be characterised as a type of omission error with slip-like properties; its occurrence is not due to deficiency at the knowledge level but at the task execution level. From early theoretical analysis to more recent empirical studies, PCE has been established as an interesting error phenomenon from a task-related or cognitive perspective. Moreover, PCE occurs in a range of different procedural tasks and can be provoked in a controlled laboratory setting. Although there are only a limited number of empirical studies on PCE, they suggest that the error is sensitive to working memory demand and is persistent in its occurrence — i.e. cannot easily be eliminated through practice alone or by the use of various motivational strategies such as being asked not to make the error. However, PCE is attenuated by dynamic presentation of visual feedback.

Recent research has shown that PCE can be induced and studied under controlled experimental environments, and this is also the approach taken by the current thesis in order to extend our understanding of PCE. The review of the literature on PCE suggests that the number of empirical studies of the error is small, and theoretical understanding of the error phenomenon is still in its infancy. Research on PCE has, traditionally, been motivated by its origin in practised procedural tasks, in which task execution involves carrying out a set of pre-defined and learned procedures. Anecdotes of PCE occurring in procedural tasks range from office equipment, such as forgetting to collect your original after photocopying, to personal machinery, such as forgetting to put the petrol cap back on after filling up petrol in your car, to public-access machines, such as forgetting to take your cash card after withdrawing money from an automatic teller machine (ATM).

The scope of the current thesis is to adopt a narrow perspective to the study of human error by focusing on error behaviour arising from an operator who has direct interaction with the artefact in question. More specifically, the current investigation adopts a psychological approach using laboratory-based experimentations. This involves designing artificial task environments to induce and study PCE in order to identify factors that provoke or mitigate the error phenomenon. The goal of this thesis is to understand PCE from a cognitive perspective; therefore, theoretical motivations and explanations are drawn from cognitive psychology.

The traditional view and empirical investigations of PCE were placed in the context of routine procedural tasks, and this is well justified since a lot of everyday examples or anecdotes of the error, as described earlier, involve carrying out procedural tasks. In order to extend our knowledge of PCE beyond its occurrence in procedural tasks, the next chapter investigates whether the prevalence of PCE extends to another type of task, namely problem-solving. Moreover, the effect of static visual cues on PCE is also examined to investigate whether the error can be mitigated.

Chapter 3

Post-completion error in problem-solving

Overview

This chapter gives the first set of empirical studies in this thesis investigating PCE using a novel methodological paradigm, namely, problem-solving. The question of whether the prevalence of PCE extends beyond procedural tasks to problem-solving is motivated by a documented error in programming practice known as the semicolon problem (Polson et al., 1992). The semicolon problem is the omission of a semicolon at the end of a coding statement, which some programming languages require. This error is essentially a PCE because inserting a semicolon at the end of a coding statement is peripheral to the main goal of constructing the piece of code in the first place. This kind of PCE resembles the second class of PCEs, mentioned in Chapter 1, which does not have an opening perturbation to the interaction but finishes with a confirmatory step. The interesting aspect about this observed PCE is that it occurs in a problem-solving domain in which the programmer is engaged in an active process of problem-solving rather than carrying out a set of learned procedures.

PCE occurring in a similar situation is also captured by the author's anecdotal experience of committing a PCE when doing arithmetic problems in primary school. The anecdote is as follows: you are carrying out some arithmetic calculations for some mathematics homework, which requires a fair level of cognitive effort. After some time working on a problem you arrive at an answer, which you believe is a correct one, then you write the answer down and move on to the next question. However, you forget to write down the unit (it could be litres, metres, miles etc.) that is required so you only get half the marks for the problem! Although an anecdote only has descriptive values, it captures some of the important characteristics of PCE: firstly, an omission of a subsidiary step (including the unit) after fulfilment of the main task goal (solving the arithmetic problem). Secondly, a considerable demand is placed on working memory during the process of solving the arithmetic problem.

Byrne and Bovair's (1997) general findings suggest that PCE is associated with situations that place a high demand on working memory. The series of experiments in the current chapter investigates the occurrence of PCE in problem-solving situations, which place considerable demands on working memory by having participants solving problems "in the head". For the purpose of the current context, a problem-solving task is taken to mean that a participant has to "figure out" the solution to a problem during task execution. In other words, the procedures of the problem task are not learned, and the problem-solver has no pre-existing knowledge of the solution before task execution. This notion of problem-solving stands in contrast to the notion of a procedural task in which task execution involves following a learned set of procedures.

The aim of the Experiment 3a was to set up an experimental paradigm employing logic problem tasks to investigate whether PCE can occur in problem-solving situations. Although the paradigm managed to generate a high level of PCE, two methodological shortcomings that possibly confound the results were identified. First, the issue of whether the participants had the correct knowledge of the PC step. Second, the problem of whether the participants had the correct knowledge of how to execute the PC step.

Experiment 3b involved correcting the methodological shortcomings from Experiment 3a redesigning the training phase of the experiment to ensure participants had acknowledged the presence of a PC step, and that they knew how to execute the action associated with the PC step. The refined methodology was successful in removing the identified problems, generating a reduced level of PCEs. The experiment also investigated the effect of a static visual cue on PCE, and the results suggest that the simple static cue present on the interface was able to mitigate the error. An unexpected finding was obtained suggesting error reduction by the static cue manipulation was more pronounced in one of the three problem tasks than in the other two. Experiment 3c was a follow-up experiment to further pursue this unexpected finding.

The series of experiments in this chapter is not an investigation of problem-solving behaviour *per se*, but rather the current experiments used problem-solving tasks as a vehicle to extend knowledge of PCE beyond our previous understanding of the phenomenon.

Experiment 3a: a preliminary study

1 Introduction

To the author's knowledge, to date, there is no empirical work that has examined the occurrence of PCE in problem-solving task, therefore, the current study is the first attempt to examine the error phenomenon occurring beyond a procedural task paradigm. The foremost objective of the current experiment was to set up an appropriate methodological paradigm to generate PCE in a laboratory setting. Further investigation into the nature of PCE is possible when the error can be reliably obtained. This in itself is a challenge to experimental studies of human error in general, as Byrne and Bovair (1997, p.42) comment that, *"For a variety of reasons, probably the most critical of which is self-monitoring on the part of the participants, laboratory studies in which such errors are produced are few and far between."*

The current experiment is not a direct replication of Byrne and Bovair's (1997) finding of PCE occurrences associated with high working memory demands, but rather examines the generality of the working memory effect in another task paradigm, namely, problem-solving. The notion of a high working memory demand is implemented in the current experiment by requiring participants to solve a set of puzzles "in the head" without the aid of any external medium, such as writing down partial solutions on paper.

The hypothesis of the current experiment is that if PCE can be reliably generated in this problem-solving paradigm, it is expected that the obtained PCE rate to exceed an error rate of 5%, which is a criterion recommended to determine the systematic occurrence of an error (Byrne & Bovair, 1997). The following section describes the problem-solving paradigm in detail.

2 Method

2.1 Tasks

The tasks adopted in this study are variants of the well-known river-crossing problems (Ernst & Newell, 1969) used in human problem-solving, reasoning and artificial intelligence research. Three puzzles were used in the current study: the Father & Son problem (FS), the Dog, Hen & Corn problem[1] (DHC) and a simplified version of the standard Missionaries & Cannibals problem ($_sMC^2$) (see Appendix A (vi-viii) for details of the problems and Appendix A (ix) for their solutions). The FS task and the sMC task are used as distractor tasks so make the participants believe the experiment is about solving puzzles but, in fact, the real purpose of the experiment is to have participants make PCE during problem-solving. The DHC problem contains a PC step, which includes the following sentence:

"*Whoever finishes using the seat always returns it to the market side, where Mr. Edison lives, so that it is easy for him to take it in each evening.*"

A PCE is operationalised, in the current context, as the omission of sending the transport vehicle back after finishing solving the problem. Goal completion of the puzzles is self-determined; the participants have to decide if they have reached the solution before carrying out the PC step. No feedback about the solutions of the problems was given until the end of the experiment.

2.2 Solving logic problems "in the head" using a text-based interface

To ensure that the task used was mentally demanding, participants were not allowed to write down any of the intermediate steps of the solution. The idea here is to have participants solve the puzzles "in the head", so they can only enter the solution steps via the computer interface developed. The interface constrains participants to work out each solution step before they type in the answer, and each answer entered is hidden away. Looking back at previous step(s) requires explicit actions at the interface.

Participants were instructed to click on the `Start` button to initiate a trial. For each step of a puzzle, participants had to type in their answers into the two text boxes and click the

[1] The use of the DHC problem, and the PC problems in Experiment 3b and 3c, was inspired by a puzzle documented in Curzon (2003).

[2] The sMC problem adopted in the current study is a simplified version of the original Missionaries & Cannibals problem, hence, the subscript "s". Instead of having three missionaries and three cannibals, the current version has only two missionaries and two cannibals. This simplified version of the problem task has a very straightforward four-step solution. This simplified version was used to keep the experimental session relatively short for the participants.

Enter button (Figure 3.1). Object(s) to be moved across the river (e.g. Fred, cat, canoe) were typed into the longer text box on the left and the sides of the river (e.g. home or market) were typed into the shorter text box on the right. Clicking the Enter button clears the text in the text boxes for the next step's entry. When participants think they have reached the solution of a problem they terminate the trial with the Finish button.

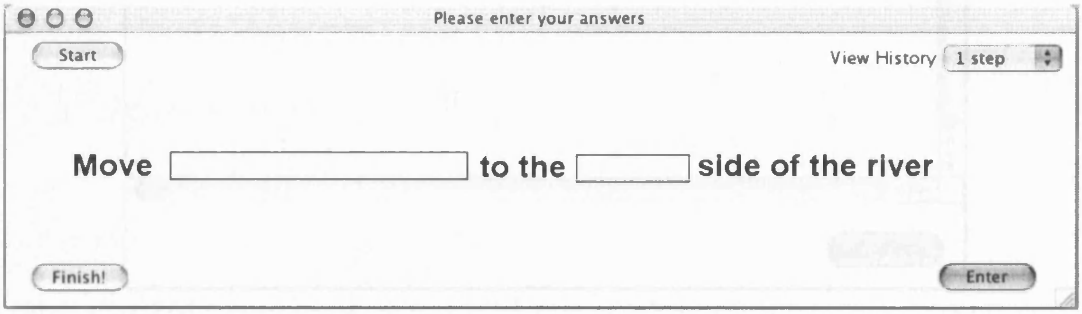


Figure 3.1: A screen shot of the main window for entering solution steps of a puzzle.

Clicking on the View History drop-down menu shows participants a record of the previous steps they have made. They could choose 1 step which only showed the previous step made, 2 steps, or All steps. Participants could access any of these three functions at any time during the experiment but they were told only to use the function when it was absolutely necessary. Figure 3.2 is a screen shot of All steps in the process of solving the sMC problem. Particpants' actions were logged by the computer.

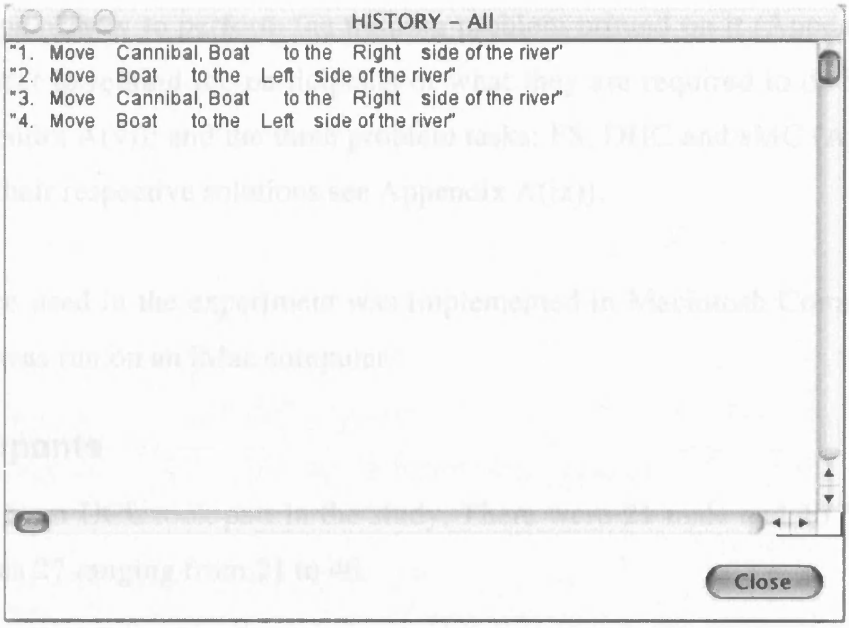


Figure 3.2: A screen shot of the `All steps` function being accessed at a particular point in time when solving the Missionary & Cannibal problem.

2.3 Design

This study is a one-factor within-subject design, with the three problems (i.e. FS, DHC and sMC) as the factor. Each problem counts as a trial; therefore, each participant received three trials. All participants were presented with all three problems in the order of: FS, DHC then sMC.

The FS and the sMC problems act as shadow tasks in order to frame the experiment to avoid participants over attending to any of the puzzles. This is important because participants' ability to self-monitor is, by and large, the most detrimental factor in reducing errorful behaviour in an experimental setting.

2.4 Materials

Materials included an instruction manual of how to enter solutions via the interface (Appendix A(i)); a training problem sheet describing a problem for a training trial (Appendix A(ii)); a recall sheet for writing down the objects and sides of the river that the participant remembered from the training problem (Appendix A(iii)); an instruction sheet with the steps of how to perform the training problem printed on it (Appendix A(iv)); a reminder sheet to remind the participants of what they are required to do in the testing phase (Appendix A(v)); and the three problem tasks: FS, DHC and sMC (Appendix A(vi to viii); for their respective solutions see Appendix A(ix)).

The interface used in the experiment was implemented in Macintosh Common Lisp 5.0 (MCL) and was run on an iMac computer.

2.5 Participants

36 students from UCL took part in the study. There were 21 male and 15 female. Their mean age was 27 ranging from 21 to 40.

2.6 Procedure

Firstly, the experimenter briefed the participant that he/she was required to solve a series of problems. The main objective of the investigation was not mentioned to the participants until after the experiment. The participant was then asked to read the instruction manual (Appendix A(i)) regarding the use of the computer interface and was allowed to ask questions about it.

Each participant was then given a training phase. The training problem sheet (Appendix A(ii)) was given to the participant and was taken away when read. The participant was then asked to recall the objects involved in the problem and the sides of the river by writing them down on the recall sheet (Appendix A(iii)). This is to familiarise the participant with having to remember the objects and sides of the river from the problem as part of the experimental procedure. An answer sheet (Appendix A(iv)), containing the order in which the objects should be entered, was then given to the participant. The participant was required to enter the various steps as printed on the answer sheet. This is to familiarise the participant with the use of the interface.

After the training phase, the participant was given a reminder sheet (Appendix A(v)) to read. This emphasised what were classified as "objects" in the problems to be followed, and that each problem has specific rules to follow in order to solve it correctly. The participant was then presented with the main trials in which they had to solve a total of three problems. Each trial had the exact same procedures as the training trial, i.e. firstly reading the problem on paper; secondly, recall of the objects and the sides of the river by writing them down on the recall sheet; and thirdly, to solve the problem.

The experimenter was in the same room as the participant throughout the experimental session. The experimenter was not allowed to answer any questions specific to the solutions of the problems, but was allowed to clarify the rules of the problems if asked by the participant. At the end of the experiment, the participant was debriefed and asked if he/she remembered the PC step in the DHC problem. This served as qualitative feedback in addition to the behavioural data logged by the computer.

2.7 Measure

The measure of primary interest is the number of PCEs committed in the DHC problem.

3 Results

Data from 9 participants were excluded from the analysis because 6 of them did not follow the instructions properly, and the remaining 3 terminated the experiment without completing the entire session. This resulted in 27 valid data points for analysis.

3.1 Are the observed PCEs systematic in occurrence?

A total of 21 PCEs were obtained in the DHC problem. In analysing the frequency of an error occurrence, researchers such as Payne & Squibb (1990) and Byrne & Bovair (1997) adopted a measure of systematicity, which is defined as the ratio of the observed error to the total opportunities for that error to occur. A 5% level threshold has been recommended to consider an error rate systematic (Byrne & Bovair, 1997). In the current study, each participant had one opportunity to make a PCE giving a total of 27 opportunities. Therefore, the observed PCE rate has a systematicity of about 78% (21/27), which is well above the recommended threshold level.

3.2 Issues related to the obtained PCEs

There are a number of notable findings related to the PCEs obtained that were not initially formulated as part of the main objective of the study. Firstly, in the DHC problem, there were two occasions in which the PC step was carried out by sending "Mrs. Jones", instead of the transport vessel — "seat", back to the other side of the river. The two participants who executed the move reported that they did not know how to send the transport vessel back, therefore, sent the carrier instead. These two responses were not classified as PCEs as the PC step was executed. The nature of these responses has a methodological implications for the experiment, as discussed in more detail in the next section.

Secondly, the participants were asked to recall objects involved in the problem on the recall sheet before solving it. Table 3.1 shows the number of participants who wrote down the transport vessel as one of the recalled objects, and the number of participants who typed in the transport vessel in the problem solutions.

	Transport vessel on recall sheet	Transport vessel in problem solution
FS	4	0
DHC	3	2
sMC	4	17

Table 3.1: Number of participants who recalled the transport vessel or included it in the problem solution.

There were only a few participants who wrote down the transport vessel on the recall sheet for each of the problems. There were substantially more participants who had typed in the transport vessel in the sMC solution; however, this was not observed for the FS and DHC problem. This suggests that there might be some difference in the sMC problem relating to the higher number of participants referencing the transport vessel in its problem solution but not in the other two problems.

Again these two findings have methodological implications for the current paradigm, as discussed in more detail in the next section.

4 Discussion

A high PCE rate fulfilled one of the main objectives set out by this pilot study — using a novel task to generate the robust error phenomenon. Previous studies investigated the error phenomenon within the boundary of highly familiar or proceduralised tasks, but the current study extends that boundary to problem-solving tasks, which are not necessarily familiar to the participants.

Although a high PCE rate of 78% was obtained, a number of findings suggest that the high occurrence of the error might be confounded by a number of factors. Firstly, in the DHC problem, two of the participants reported that they did not know how to send the transport vehicle back at the end of the solution; therefore, the carrier was sent back instead to fulfil the PC step. Although this response was only observed from two participants, it suggests an important methodological implication for the current paradigm — the need to ensure participants have the correct knowledge of how to execute the PC step.

A further finding relating to the knowledge of the PC step execution is the failure to include the transport vessel in the intermediate steps of the DHC problem solution. 25 out of the 27 participants did not type in the transport vessel in the problem solution, even though it was explicitly stated in the instructions that everything that gets moved across a river ("a canoe" was provided as an example) is classified as an object and should be included in the solution. Moreover, only three participants recognised the transport vessel as an object when asked to write down on the recall sheet. On this basis, the cause of the high PCE rate is unclear: whether participants had the correct knowledge of how to execute the PC step but genuinely forgot about the required step; or they simply did not know one could move the transport vessel across the river on its own and, therefore, did not perform the required action.

However, the same finding was not observed in the sMC problem. There were a relatively higher number of participants (17 out of 27) who included the transport vessel in the solution steps. A closer examination of the sMC problem suggests that this is most likely due to a peculiarity that is not present in the FS and DHC problem. In the sMC problem, there are steps in which the transport vessel has to be sent back on its own without any passenger on it in order to proceed further in the solution. This feature in the problem most likely had prompted the participants about the permissible action of sending the transport vessel alone.

The second methodological concern is related to whether participants had encoded the PC step as a required step to the problem solution. Participants were asked about their recollection of the PC step in the DHC problem at the end of the experiment. 8 out of 27 participants responded that they were aware of the PC step but had concentrated on the main goal and forgotten about it. This anecdotal evidence plus some of the behavioural data (6 participants who did *not* commit the PCE) suggests that 14 out of 27 participants had, more or less, acknowledged the PC step as an integral step to the solution. However, this is mostly anecdotal evidence, and the data also suggests that the rest of the participants were not aware of the PC step as part of the solution.

Finally, the following is not a methodological issue but an issue concerned with the problem-solving nature of the currently adopted task. One might argue against the definitional issue over the error and, consequently, the suitability of the task for investigating the phenomenon. This is because in the current problem-solving situation, not every participant was able to solve the problem correctly. The traditional view of PCE (e.g. Byrne & Bovair, 1997) is an error, which "*people have the knowledge required to perform the task correctly, but still fail on occasion.*" (p. 32). The direction of investigation taken by this view led one to look at the error in well-practiced routine activities. However, the very essence of PCE resides in a particular task characteristic, in which a step needs to be carried out *after* completing the main goal. Since goal completion in the DHC task is self-determined, the importance lies in the fact that the participants think they have completed the goal more than the "correctness" of their solutions. Once the participants have decided the main goal has been fulfilled, the remaining step in the task, by definition, is a PC or "clean-up" step.

Nevertheless, the methodological concerns mentioned earlier suggest that it is important to ensure participants have correct knowledge of the PC step and also the knowledge of how to execute the step. The following are a number of possible refinements to rectify the identified methodological shortcomings.

Firstly, the training trial may be modified to include an example problem resembling closely the current sMC problem. The important point is to show the participants that the transport vessel can be moved on its own, to ensure their having the knowledge of how to execute the permissible action. Secondly, in order to ensure participants have acknowledged the PC step as a required step, a questionnaire could be given to participants before solving a problem. Questions such as "Is it true or false that the hen can not be left alone with the corn?", "Is it true or false that the seat is to be returned to the market side of the river after use?", etc. would provide a way to check that the participants have encoded the PC step as part of the solution. Furthermore, in order to produce a high enough error rate for subsequent studies, each participant could be given more than one PC problem. Variations of the current PC problem could be used to provide more error opportunities for each participant.

Finally, a manipulation that could be tested in the subsequent experiment is the effectiveness of a simple cue on PCE. This idea comes from one of the observations from the current study: in the DHC problem, the transport vessel ("seat") could be regarded as a cue related to the PC step (returning the "seat" to the market side) of the task. The responses given by the participants who did not include the word "seat" might be related to the high PCE rate obtained. As discussed earlier, although the failure to include the transport vessel in the solution may have been problematic to the validity of the obtained high error rate, the error rate might also be related to the absence of relevant visual cues. Manipulating the presence of the transport vessel during the problem-solving process allows one to investigate whether PCE could be mitigated by the presence of a cue relating to the PC step. Nevertheless, before the manipulation of visual cues can be investigated, the identified methodological shortcomings should be rectified to ensure the validity of the error rate obtained. This is the motivation for the next experiment.

Experiment 3b: refined methodology and the effect of static visual cue

1 Introduction

There are two aims in this experiment, the first aim is to eliminate the identified confounding factors by refining the experimental methodology. The major refinements made in this experiment include the following: firstly, an instructional phase with a real problem-solving example (rather than an abstract problem as used in the previous experiment) was used. The example not only illustrated how to use the interface for entering answers but how the transport vessel may be moved on its own when necessary. Secondly, in the training phase, a problem with a PC step was used (the structure of the problem closely resembles the sMC problem from the previous experiment). A questionnaire was also used to assess whether participants were aware of the required PC step. Finally, the testing phase involves three PC problems and three non-PC problems acting as shadow problems. It is expected that the overall rate of PCEs would decrease in this current experiment as a result of implementing the methodological refinements. The refined methodology is described in more detail in the "Method" section. It is hypothesised that there will be a decrease of the overall PCE rate but the error rate will still be above the 5% systematicity criterion.

The second aim of the current experiment is to investigate the effect of static visual cues on PCE. In a recent study, Chung and Byrne (2004) examined the effect of visual cues on reducing PC errors. The main PC task used in their experiment was the procedural task (i.e. Phaser task) used by Byrne and Bovair (1997). The experiment tested two types of reminders; a mode indicator relying on static contextual information and a dynamic cue occurring just before the PC step. It was found that the visual cue with dynamic movement (flashing arrows in this case) was effective in guarding against the error, but the static indicator was not; this is consistent with Lee's (1992) findings. Furthermore, it was found that, to be effective, a cue should be meaningful in relation to the PC action, and occur just before the PC step. However, there are examples of situations where the artefact in use might not necessarily be able to track one's state in a task; for example, when programming a VCR, there is no way the VCR could tell when the user has finished the programming task. In such situations, the implementation of a just-in-time cue will not be possible and other means of cueing should be adopted.

No existing theoretical accounts support predictions about the effect of static visual cues on PCE in problem-solving tasks; therefore, the nature of this investigation is exploratory. The second hypothesis of the current experiment is that the static visual cues should be effective in mitigating PCE in the current problem-solving tasks, because of the simplicity of the interfaces used in the current experiment. The basis of this hypothesis is that there are fewer visual distractions in simpler interfaces and this would make a static visual cue relatively more salient than in more complex interfaces, such as the one of the Phaser task. The manipulation of static visual cues in the current study involves using two different interfaces to solve the problem tasks. These two interfaces are described in detail in the following section.

2 Solving logic problems "in the head"

Logic problems such as the Missionaries & Cannibals problem, and similar river-crossing problems, were used in this study to provoke the occurrence of PCEs. The idea of using logic problems, as opposed to using routine procedural tasks like Byrne & Bovair's (1997), is thought to be a novel experimental paradigm in investigating PCEs. Furthermore, solving these logic problems is mentally demanding in terms of working memory resources when one is not allowed to write down the intermediate steps of the solution. Therefore, based on Byrne & Boviar's finding that increasing working memory demands increases the chances of making PCEs, it was thought that solving logic problems "in the head" would be an effective way of producing PCEs without using a concurrent secondary task.

2.1 Two interfaces — text-based and menu-based

In order that the participants to solve the problems "in the head", two interfaces were designed for the two experimental conditions: the Text (Txt) condition and the Pop-up (Pop1) condition. Common to both interfaces is that each step of the solution entered is hidden away: looking back at previous step(s) requires explicit actions on the interface.

In the Txt condition, participants have to type in their answers into the two text boxes and click the `Enter` button (Figure 3.3). Object(s) to be moved across the river (e.g. Mrs. Jones and Hen) are typed into the longer text box on the left and the sides of the river (e.g. Home or Market) are typed into the shorter text box on the right. Clicking the `Enter` button clears the text in the text boxes for the next step's entry. If a mistake is made the participant can press the `Reset` button and start from the beginning. When the participant thinks he/she had reached the solution of a problem they can terminate the trial with the `Finish` button.

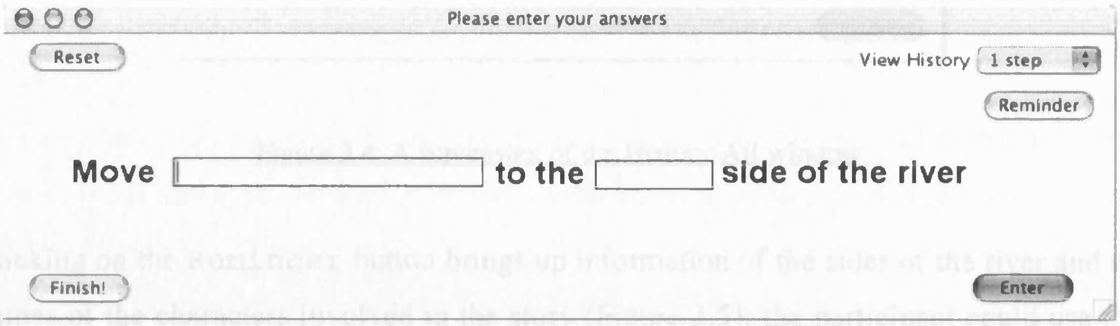


Figure 3.3: A screenshot of the Txt interface.

Clicking on the `View History` drop-down menu shows a record of the previous steps they have made. They can choose `1 step` which only shows the previous step made, 2 `steps`, or `All steps`. The participant can access any of these three functions at any time during the experiment but they were told only to use the function when it was absolutely necessary. Figure 3.4 is a screen shot of `All steps` in the process of solving the DHC problem.

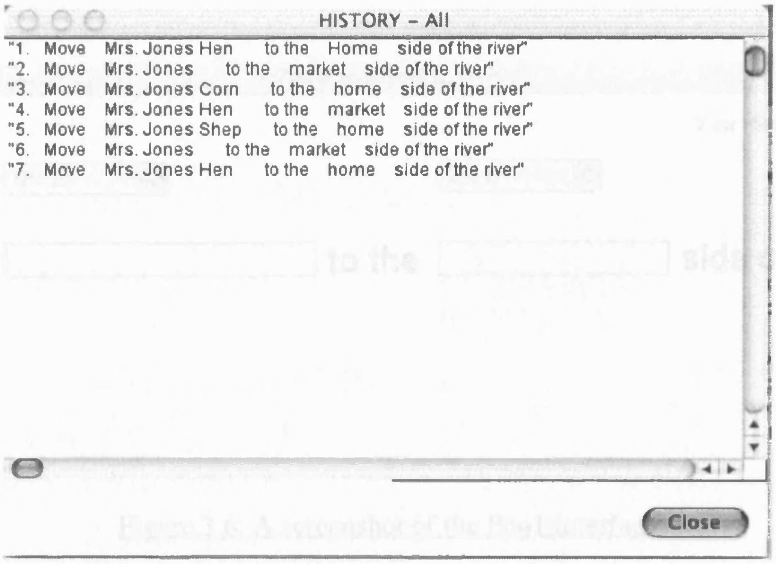


Figure 3.4: A screenshot of the History-All window.

Clicking on the `Reminder` button brings up information of the sides of the river and the names of the characters involved in the story (Figure 3.5); the participant could use this function when he/she forgot any of the information but was instructed to keep the usage to a minimum.

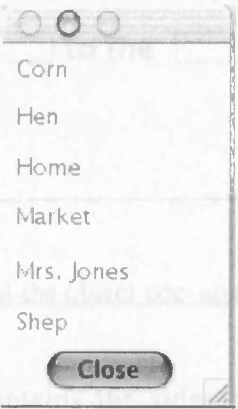


Figure 3.5: A screenshot of the Reminder window.

2.1.1 Control of the visual cue for the post-completion step

In condition Pop1, participants entered their answers into the text boxes by selecting the appropriate item(s) from two pop-up menus (Figure 3.6).

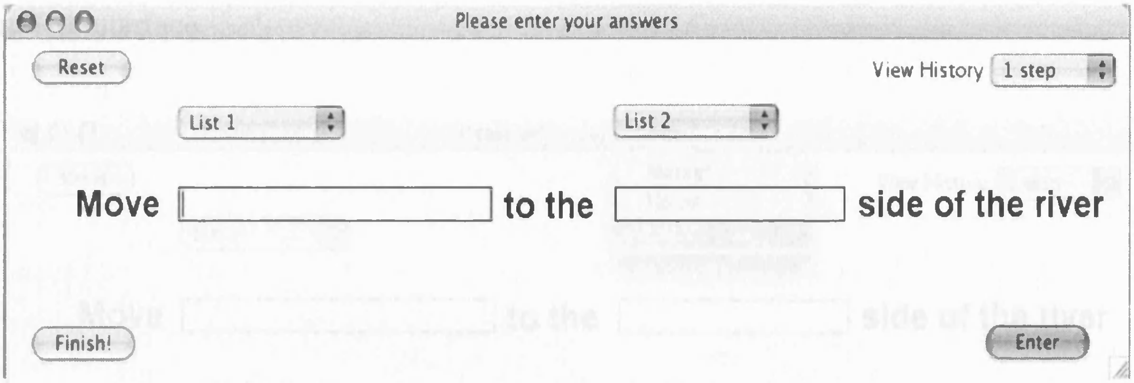


Figure 3.6: A screenshot of the Pop1 interface.

The pop-up menu on the left contains a list of object(s) to be moved across the river, including the transport vessel (i.e. "seat" in the DHC problem) used in the story (Figure 3.7).

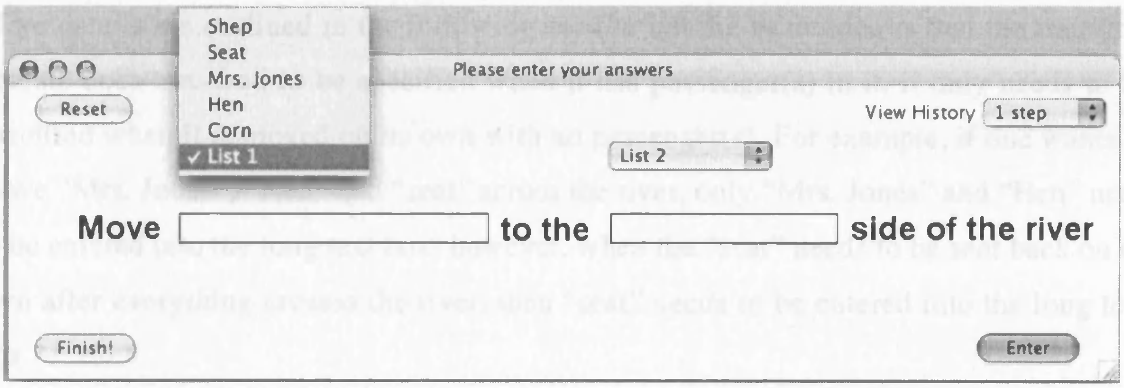


Figure 3.7: A screenshot of the object pop-up menu in the Pop1 interface.

The pop-up menu on the right contains the sides of the river (Figure 3.8). The main difference, in terms of the information available to the participants during the problem-solving process, between the Pop1 and the Txt interface is the presence of the transport vehicle from the problem story. The transport vehicle itself is thought to be a visual cue reminding one of the post-completion step of sending it back to the other side of the river at the end. The `Reminder` window in the Txt interface contains all of the information, but the transport vehicle, provided in the pop-up menus in the Pop1 interface. This is to ensure there are no transport vehicles, which might act as a source of visual reminder, in the Txt interface.

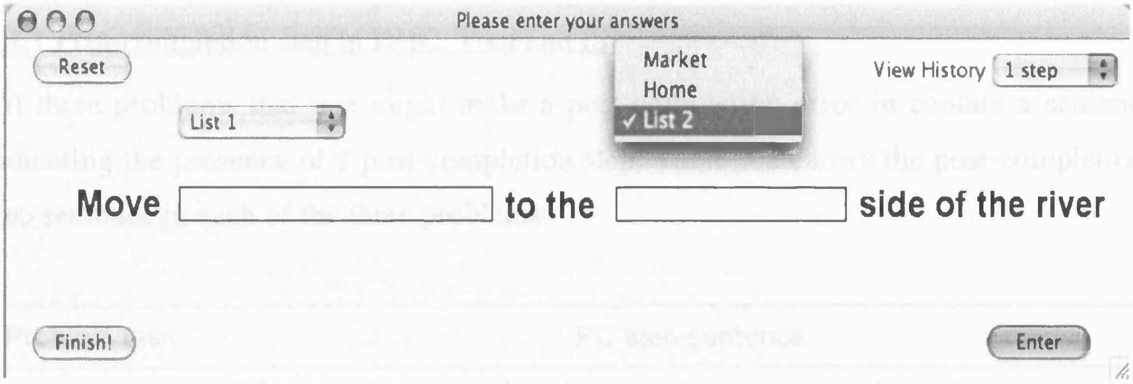


Figure 3.8: A screenshot of the river-side pop-up menu in the Pop1 interface.

It is also worth mentioning that the control of visual cueing of the PC step between the two conditions is done through explicit instructions in the instruction and training phase. More details are outlined in the following section but the basic idea is that the transport vehicle does not need to be specified when it has passenger(s) in it, it only needs to be specified when it is moved on its own with no passenger(s). For example, if one wants to move "Mrs. Jones", "Hen" and "seat" across the river, only "Mrs. Jones" and "Hen" need to be entered into the long text box; however, when the "seat" needs to be sent back on its own after everything crosses the river, then "seat" needs to be entered into the long text box.

3 Method

3.1 Tasks

As in Experiment 3a, variants of river-crossing problems (Ernst & Newell, 1969) were used in this experiment. Six problems were used: the Father & Son problem (FS), the Three Guests problem (3G), the Lion, Monkey & Banana problem (LMB), the Dog, Hen & Corn problem (DHC), the Torch problem (Trch) and the Itchy & Scratchy problem (IS). The latter three problems were modified to contain post-completion steps which required one to send the transport vehicle back to the other side of the river as the very last step of the solution. Each problem is adapted into the form of a story so that the post-completion step is coherent with the story (see Appendicees A(vi & vii) and B(iii to vi) for details and Appendices A(ix) and B(viii to xi) solutions for each of the problems).

3.1.1 Post-completion step in DHC, Trch and IS

All three problems that one might make a post-completion error in contain a sentence indicating the presence of a post-completion step. Table 3.2 shows the post-completion-step sentence in each of the three problems:

Problem task	PC step sentence
DHC	Whoever finishes using the seat always returns it to the market side, where Mr. Edison lives, so that it is easy for him to take it in each evening.
Trch	When they reach the other side of the river, they have to send the boat back to the forest so that no one will notice it has been used
IS	They realise that they have to send the crate back to the initial side when they reach the other side so no one would suspect that it had been used

Table 3.2: The sentences used in indicating a post-completion step in the DHC, Trch and IS tasks.

3.2 Design

A between-subject design was used in this study. The between-subject manipulation was the Txt or the Pop1 interface used in solving the problems. All participants did all six problems using either one of the interfaces. The post-completion problems were placed in positions 2, 4 and 6 in the sequence of problem presentation, and the sequence was counterbalanced using a 3×3 Latin Square. Positions 1, 3 and 5 in the problem presentation sequence were filled with non-post-completion problems and were counterbalanced again using a 3×3 Latin Square.

3.3 Materials

Materials included a Power Point presentation, which described the instructions involved in using the interface; a training problem sheet (Appendix B(i)) with an example problem task printed on it; a check list (Appendix B(ii)) which contains 14 statements relating to the training problem; and the six problem tasks printed on paper (Appendices A(vi & vii) and B(iii to vi)).

The two interfaces used in the conditions Txt and Pop1 were implemented in MCL 5.0 and they were run on an iMac computer. The Power Point presentation for the instruction phase was also run on the same computer.

3.4 Participants

64 participants participated in the current study; the majority of which were UCL students. There were 31 males and 33 females, age ranging from 18 to 50 with a mean of 25.6. Each participant was paid £3 for participating in the experiment; moreover, as an incentive to encourage performance an extra 50p was paid for every correctly solved problem. Therefore, each participant could get paid a maximum of £6.

3.5 Procedure

All participants went through the following procedures:

3.5.1 Instruction phase

First, in the instruction phase each participant was presented with a set of instructions (in Microsoft PowerPoint format) about solving a simple river-crossing example problem via the specially designed interface. The sets of instructions were identical in the Txt and the Pop1 condition except for the details about the computer interface involved for entering answers. The solution of the example problem was walked through during the instruction presentation; the first step of the solution was an explicit action of moving the transport vehicle to one side of the river. The purpose of this very first step was to ensure the participant knew one could enter the transport vehicle alone into the left text box on the interface. Furthermore, the instruction also stated that the transport vehicle only needed to be specified when it is moved on its own with no passenger(s) in it. This is to eliminate entering the transport vehicle into every step of a solution to avoid confounding the absence of post-completion reminders in the Txt condition.

The participants were also told that for every problem solved successfully they would receive a payment of 50p. The monetary incentive was designed to encourage participants to fully engage with the problems and complete them as best as they could. If participants were not spending any effort in thinking of the solutions, they could just give the wrong answers very quickly.

3.5.2 Training trial

Second, after the instruction phase, participants were asked to perform a training trial which consisted of reading a simple river-crossing problem (which was printed on paper, see Appendix B(i) for the problem story), then it was taken away from the participant so that he/she had to memorise the rules, names of the object(s) to be moved and the sides of the river from the problem. To ensure the participants knew of the presence of a post-completion step in the problem, a reminder ("Please also send the boat back to the coffee shop when finish taking everything across") was placed at the end of the story in addition to the post-completion sentence ("the boat is always sent back to the coffee shop side of the river, as a convenience to the staff for the next order ").

Before the participants started entering their answers into the interface, they were asked to complete a check list (Appendix B(ii)) indicating whether they had acknowledged the post-completion step as a required step to the solution. The check list consisted of 14 statements (6 true, 8 false) which the participants had to identify as true or false. The true statements were also required to be rated as how important (Not at all, Moderate or Very) they would be in solving the problem. The statement "The boat is always sent back to the coffee shop side of the river after use" was used to assess the acknowledgement of the post-completion step.

Feedback on any errors made (including PCE) was given to the participants at the end of the training trial so that they knew if they had completed the trial successfully. It is worthwhile to mention that the solution to the training problem (Appendix B(vii)) involved the action of explicitly sending the transport vehicle to the other side of the river in intervening steps. The rationale is to further make sure participants know the permissible action of sending the transport vehicle alone to one side of the river.

3.5.3 Test trials

Third, following the training trial, participants were asked to perform the test trials which consisted of a series of six logic problems (Appendices A(vi & vii) and B(iii to vi)). As in the training trial, participants were asked to read a story and enter their answers into the interface on a computer. A check list was not required in the test trials phase to eliminate the possibility of reminding them about the post-completion steps in the three post-completion problems. Participants were also instructed to hand the problem story, which was printed on paper, back to the experimenter as soon as they finished reading it. The participants were then asked to start solving the problem by entering the answers into the interface on a computer.

Participants were told that there would be no time pressure in solving the problems; however, they were told that the session would last for about an hour and that the experimenter would indicate if they should move on to the next problem when they could not work out the solution to a problem. When the participant indicated that they could not arrive at the solution or they had spent too long on a particular problem, the experimenter would ask him/her to give a second best answer (which would depend on the context of the problem). The rationale was to have participants finishing the problem regardless of the correctness of the solution and see if they would still commit post-completion errors in the three problems.

No feedback to the solutions was given during the test trials and this was to minimise the effect of explicitly reminding participants about the post-completion steps in the three post-completion problems. Feedback to the solutions was given at the end of the experiment. During an experimental session, the experimenter was present in the same room as the participant. The experimenter was not allowed to answer questions regarding the solutions to the problems but was allowed to clarify questions about the rules of the problems. At the end of the experiment, participants were debriefed about the objective of the experiment. Each session lasted approximately an hour.

3.6 Measures

The measure of primary interest is the number of PCEs. Measures of secondary interest such as completion times of the problem tasks, usage of the History and Reminder functions were also collected.

4 Results

Data from five participants were excluded because they withdrew from the experiment session after the first trial. A total of 59 participants were included in the data analysis; 29 in the Txt condition and 30 in the Pop1 condition.

The following analyses were carried out on data concentrating on the three post-completion problem tasks. Since each participant had three possible opportunities of making at least one PCE, the entire data set should have yielded a total of 177 data points; 87 in Txt condition and 90 in Pop1 condition.

12 data points were excluded from the analyses resulting in a total of 163. Five of the excluded data points were missing data: trials not completed by the participants because they could not give the second most approximate solution when prompted to do so. Seven of the data points were either indeterminate as PCEs: participants reported that he/she was unclear about the PC step before finishing the trial; or contained solutions that did not follow the instructions of the problem: for example, sending the vehicle on its own at intermediate steps, mention of items that should not be mentioned, sending more items than the vehicle can hold, etc.

The 12 excluded data points came from different participants, but other data points from these individuals were intact. Missing or excluded data yielded a total of 23 and 24 valid cases from conditions Txt and Pop1 respectively for data analysis.

4.1 Responses from checklist in the training trial

Almost all participants (43 out of 47) rated the PC step in the checklist as "very important". Three participants gave an "important" response and one participant did not give the item any rating. The responses from the checklist suggest that the PC step, as indicated by the PC sentence and the reminder sentence, was acknowledged as part of the integral solution in the training trial.

4.2 PCE rates in Txt and Pop1

Table 3.3 summarises the number of PCEs in the two experimental conditions.

	Txt	Pop1
No. of PCEs	33	16
No. of PCE opportunities	69	72
Error rate (%)	47%	22%

Table 3.3: Summary of number of PCEs obtained in the two experimental conditions.

Overall, more than half of the participants (29 out of 47) committed at least one PCE. Collapsed across problem tasks and conditions, a total of 49 PCEs were generated out of 141 opportunities (each participant had three PCE opportunities). Therefore, the overall PCE rate (no. of errors divided by no. of opportunities for that error) is about 35%. This result supports the first hypothesis of the current experiment that the occurrence of PCE in these problem-solving tasks is systematic because the obtained error rate exceeds the 5% systematicity level.

4.3 Difference in PCE occurrences between Txt and Pop1 condition

A score was calculated for each participant based on the number of PCEs committed, with one PCE counting as a score of 1; so each participant could have a minimum score of 0 and a maximum of 3. Table 3.4 shows the respective mean, median and interquartile range of the PCE scores.

	Txt	Pop1
Mean	1.43	0.67
Median	2.00	0.50
Interquartile range	0.00 – 2.00	0.00 – 1.00

Table 3.4: The means, medians and interquartile ranges for the two conditions.

A Mann-Whitney U test was used to examine if there was a significant difference in the scores between the Txt condition and the Pop1 condition. The difference was found to be significant, U = 164 (N_1 = 23, N_2 = 24), p = .012, suggesting that there were reliably fewer PCEs in condition Pop1 than condition Txt. This difference supports the second hypothesis that PCE can be mitigated through the use of a static visual cue.

4.4 PCE occurrences in each individual problem task

While the above results suggest a significant overall effect of error reduction across the three PC problems, the effectiveness of the Pop1 interface is expected to be present in each individual PC problem tasks. The following analysis was used to examine whether the error reduction effect is general to all problem tasks or only specific to certain problem tasks.

Table 3.5 shows the respective number of PCEs for all three PC problem tasks (DHC, Trch and IS) in both conditions.

	Txt			Pop1		
	DHC	Trch	IS	DHC	Trch	IS
No. of PCE	10	9	14	5	6	5
No. of PCE opportunities	23	23	23	24	24	24
Error rate (%)	43.5%	39.1%	60.9%	20.8%	25%	20.8%

Table 3.5: No. of PCEs in each individual problem task in condition Txt and Pop1.

In each PC problem task, each participant either made the error or not. Therefore, Chi-Square comparisons were made for each problem task *separately* to assess any significant differences in the number of PCEs between Txt and Pop1 conditions. Post hoc comparisons using Bonferroni correction obtained a significant difference in PCE in the IS task, χ^2 = 7.817 (1, N = 47), p = .005. In descriptive terms, the PCE rate in condition Txt was about three times the PCE rate in condition Pop1. However, no significant difference was found in the DHC, χ^2 = 2.77 (1, N = 47), n.s., and Trch problem task χ^2 = 1.08 (1, N = 47), n.s..

4.5 Difference in number of PCEs between the three problem tasks

The next logical question to ask is whether the significant error reduction effect in the IS problem task is due to a higher PCE rate inherent to the problem itself. PCE rates are compared across the three problem tasks within condition Txt and condition Pop1. Cochran Q test was used to assess differences between more than two within-subject dichotomous variables. Although Table 3.5 shows that, in condition Txt, there was a higher number of PCE occurrences in IS followed by DHC then Trch, Cochran Q test yielded no significant difference between the three problems in terms of the number of PCEs, $Q = 3.231$ ($df = 2$), n.s.. Differences in the number of PCEs between the three problem tasks in condition Pop1 was also found to be not significant, $Q = .167$ ($df = 2$), n.s..

5 Discussion

The current experiment obtained an overall PCE rate of 35% across the two experimental conditions and this supports the first hypothesis that PCE can occur systematically in a problem-solving paradigm. In Experiment 3a, which used the same paradigm as the current experiment but with only one PC problem task (DHC), the obtained PCE rate was 80%. The high rate of PCE was thought to be associated with a couple of methodological shortcomings. In the current experiment, the obtained PCE rate in the DHC task was 43.5% in the Txt condition alone. The still relatively high error rate, but decreased from the previous experiment, is an anticipated and positive finding. The decrease in the error rate is taken to indicate that the current experiment (the Txt condition specifically) has successfully "debugged" the shortcomings of the previous experiment; through the refined instructions and training trials to ensure participants had knowledge of the PC step present in the problem and had knowledge of how to execute the PC step. This is evident in the responses given by all, but one, participant rating the PC step as either "very important" or "important" to the overall solution of the problem.

The overall PCE rate was found to be significantly different between the Txt and the Pop1 condition; there were fewer PCEs in the Pop1 condition than the Txt condition. The current result suggests that, regardless of the different problem tasks, PCEs were less likely to occur when the Pop1 interface, containing a visual cue related to the PC step, was used to solve the PC problem tasks. Finer data examination at the individual problem task level revealed significant error reduction between the two interface conditions in the IS problem task; the magnitude of the difference in PCE occurrences suggests that the error is about three times as likely to occur in Txt as in Pop1. Although the same error reduction pattern was not found to be statistically significant in DHC or Trch, the trend of the data is consistent with the overall finding that fewer PCEs occurred in Pop1 condition. The overall finding suggests that the use of a simple static visual cue could mitigate PCE occurrences, and the error reduction effect was relatively more substantial in one of problem tasks – IS. This finding provides support to the second hypothesis of the current study that PCE can be mitigated through the use of static visual cues.

The differential error reduction effect between different problem tasks was not an anticipated outcome. One possible reason for this could be the problem tasks themselves. This speculation is made on the basis that the three problem tasks are not isomorphic. Since Byrne & Bovair's (1997) findings suggest that high working memory load is a major determinant on the occurrences of PCE, it is not unreasonable to speculate that the different problems used in the current study might impose different working memory demands on participants and, consequently, result in different reduction rate of PCE occurrences between the two different interfaces used. To pursue this speculation, the next section is a qualitative analysis of the three problem tasks.

5.1 Analysis on the three PC problems (DHC, Trch & IS)

Problem-solving research work has been carried out to determine how difficult a problem is by quantifying its imposed working memory load; for example, Kotovsky & Kushmerick (1991) compared participants' performances on isomorphic versions of the Tower of Hanoi problem and developed a computational model to explain the differences in working memory load imposed by the isomorphs. However, it is difficult to quantify the level of difficulties in the problems used in the current study and this is, largely, because the three problems used are not isomorphic, i.e. they differ in terms of features in their problem space structures. As a result, in any subsequent comparisons between the three problems it is impractical to pinpoint exactly which dimensions one problem is more difficult than others. However, by adopting some of the dimensions in which problems are analysed in the problem-solving domain, the objective of the following analysis is to outline some of the main qualitative differences between the DHC, Trch and IS problem task in order to examine if the differences in PCE reduction rate are related to inherent characteristics of the problems.

The first dimension to look at is performance level in terms of the number of correct solutions for each problem task across the two experimental conditions. This dimension might provide a rough indication of the relative difficulty of the problems; the more participants able to solve the problem, the less difficult it is. Looking at the data on a descriptive level, the DHC problem was solved correctly by 70% (33 out of 47) of the participants, the Trch problem was solved correctly by 15% (7 out of 47) of the participants, and the IS problem was solved correctly by 23% (11 out of 47) of the participants. The differences in the number of correct solutions arrived at in each problem suggest that DHC might be the least difficult in the problem set; however, it is not clear which problem is more difficult among Trch and IS based on this performance level dimension alone.

Other measures such as completion times and usage of the History and Reminder functions have been taken for each problem. Table 3.6 shows the mean completion times for the three problem tasks in both conditions. The general trend suggests that IS took the longest time to complete followed by Trch and then DHC. The difference in mean completion times between condition Txt and Pop1 is not more than 60 seconds for any problem. IS appears to have taken the longest time to complete. However, this measure alone is not informative enough to differentiate how difficult or demanding in terms of working memory the problem tasks are. The completion times may well be a by-product of the number of steps required to solve the problem, the more steps a complete solution requires the more time may be required to complete it.

	Txt	Pop1
DHC	197s (101)	241s (172)
Trch	341s (275)	325s (219)
IS	616s (283)	564s (248)

Table 3.6: Mean completion time and standard deviation (in bracket) for each PC problem task in the two experimental conditions.

Table 3.7 summarises the means and standard deviations of the number of History function usage (to recall previous steps carried out).

	Txt	Pop1
DHC	0.65 (0.98)	0.83 (1.74)
Trch	0.74 (1.49)	0.96 (0.95)
IS	2.5 (2.7)	2.71 (2.26)

Table 3.7: Mean and standard deviation (in bracket) of History functions usage for each PC problem task in the two conditions.

Differences in mean usage between the two conditions for each problem task are not more than 0.3. IS had the most usage followed by Trch then DHC in the History functions. This measure might indicate a relationship with demands on working memory; the higher the working memory demands the more the usage required. However, it is not clear how the minimal difference in the functions usage between DHC and Trch might agree with the rather substantial difference between the two problems in terms of correct solutions achieved.

The usage of the Reminder function might be related to the number of entities involved in a problem task; the more entities, the more the usage required. Such a relation might, arguably, add extra demands on working memory. The Reminder function is only present in the Txt condition and was used most often in the Trch problem ($M = 2.35$, $SD = 2.14$) followed by the DHC problem ($M = 1.57$, $SD = 1.8$), and followed by the IS problem (M

= 1.4, SD = 1.4). However, the pattern of data on reminder usage is not consistent with the pattern of the performance level: the Trch problem had the fewest correct solutions, followed by the IS problem and then the DHC problem.

By drawing out the problem space graph for each individual problem (see Appendix C) one can then start describing the differences among the problems in terms of other dimensions. The diagrammatic representations of these problem space graphs are adapted from Jeffries, Polson, Razran & Atwood's (1977) problem space graph for the Missionaries-Cannibals problem (p. 414). Following some of the dimensions discussed as being important in determining the difficulty of a problem in problem-solving research (e.g. Kotovsky, Hayes & Simon, 1985), the following dimensions have been chosen to assess the relative difficulty in the current problem set: a) minimum solution path: the least number of moves involved in arriving at the solution, b) number of "moving away from goal" steps: the number of moves that result in a state which apparently removes one further away from the goal state, c) number of branches: the number of branches in the problem space that leads to an illegal state, and d) number of entities involved to test the legality of a move. Table 3.8 summarises the three problems in terms of these four dimensions.

	Problem task		
	DHC	Trch	IS
Minimum solution path	9	7	13
No. of "moving away from the goal" step	1	0	1
No. of branches	7	N/A	18
No. of entities involved to test the legality of a move	4	4	6

Table 3.8: The three problems (DHC, IS and Trch) and their corresponding characteristics in terms of four chosen dimensions.

First of all, let us consider the comparison between DHC and IS. These two problems are similar in that they both include a "moving away from the goal" step in their solution paths; it is this step that most participants had problems with (Thomas, 1974 & Greeno, 1974). Furthermore, the river-crossing paradigms in DHC and IS are similar in that each state is reversible to its previous state. The differences between the two problems are that DHC has a shorter minimum solution path, fewer branches and fewer entities to consider when testing the legality of a move. Differences in these dimensions might make the DHC problem less difficult than the IS problem, accounting for why more participants managed to solve it correctly.

However, the same logical comparison does not apply to Trch as it has the shortest minimum solution path, no "moving away from goal" step and the same number of entities as DHC, but it has the smallest number of participants solving it correctly. A closer look at the problem space of Trch suggests that it has a couple of inherently different characteristics to the other two problems.

Firstly, the river-crossing paradigm in Trch is different to the others in that each state is *not* reversible to its previous state because of the time component in the problem; every move made in the problem adds extra time to the solution path. Secondly, the type of branches in Trch is different in that a branch from the start state leading to an illegal state is not immediately apparent, i.e. they are like "traps". Once one makes an incorrect move from the start state, the illegal state does not appear until a few moves later, whereas an illegal state is immediately apparent usually after making one single move in DHC or IS. Therefore, it is impractical to quantify the number of branches in the Trch problem task, which is indicated as N/A in table 3.8.

The presence of "traps" in the Trch problem might impose a relatively larger exploratory space on participants compared to the other two problems and this might explain the lower performance level in the Trch problem even though it has, for example, a shorter minimum solution path than DHC. The relatively low performance levels in Trch and IS might suggest that they are, more or less, of equal difficulty but for different reasons. Although IS does not have the "traps" property of Trch, the difficulty of the problem might be because of its longer minimum solution path and more entities involved in testing the legality of a move than the Trch problem.

The performance level data together with the analysis outlined above suggest that, among the problem set used in the current experiment, the DHC problem is the least difficult and the Trch and the IS problem are, more or less, equally difficult. Although demands on working memory imposed by the problems might be different and might correlate with level of difficulty, only qualitative comparisons are possible and informative since the three problems adopted are not isomorphic.

With respect to the relatively more substantial error reduction found in IS: if the effect of error reduction was related to the difficulty of the problems, then one should expect to see Trch and IS having the same pattern of results and the DHC problem having a reversed result pattern. However, this hypothetical pattern of results was not supported by the data and this rules out the speculation that the differential error reduction rate in the three problems being related to the differences in difficulty or the qualitative characteristics inherent in the problems themselves.

5.2 Pop-up menus in the Pop1 condition

A second speculation on the differential error reduction is made on the basis that the pop-up menus used for each of the three problem tasks exhibit some differences at a physical level. Figure 3.9 are the screen shots of the three pop-up menus used in Pop1 condition for DHC, Trch and IS.

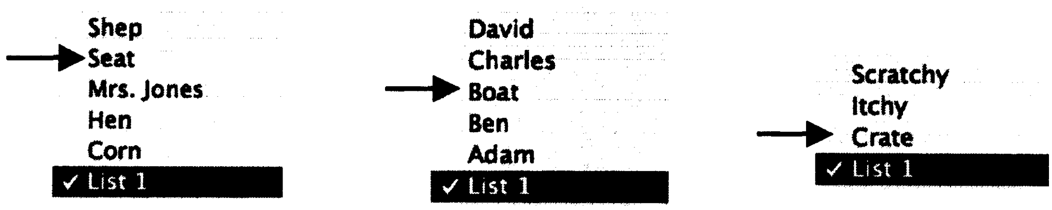


Figure 3.9: Pop-up menus used in Pop1 condition for DHC (left), Trch (middle) and IS (right). Arrows indicating the PC cues.

One common characteristic of the DHC and Trch menus is that there were five items in each of the menu lists, whereas there were only three items in the IS menu list. In addition, the PC cue is embedded within the list in the DHC and Trch cases, but appears at the bottom (right next to the initially highlighted item) in the IS case. When using these lists, participants' visual attention on the PC cue (i.e. Seat or Boat) might have been distracted by other items in the list for DHC and Trch, making the cue less conspicuous and, consequently, less effective as a reminder to the PC step when compared to the IS list.

The following experiment is designed to examine whether the differential error reduction among the three problem tasks is related to the differences in the cue position of the menus.

Experiment 3c: a follow-up study

1 Introduction

This experiment is a follow-up from Experiment 3b to examine the effect of cue position in pop-up menus on PCE occurrences. More specifically, by controlling for the relative position of the PC cue in the pop-up menus, this experiment is set out to examine if the differential error reduction observed in the previous experiment still persists.

1.1 Modification of the pop-up menus

A modification was made to each of the three pop-up menus used in Experiment 3b. Figure 3.10 shows the screen shots of the modified pop-up menus for the three problem tasks.

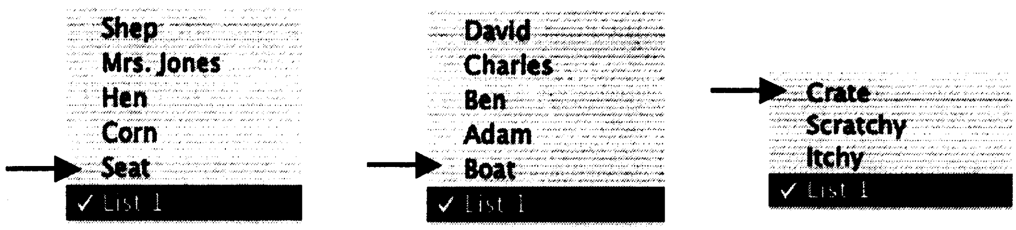


Figure 3.10: Modified pop-up menus for DHC (left), Trch (middle) and IS (right). Arrows indicating the PC cues.

The lack of reduction of PCEs in the DHC and Trch problem task was attributed to the PC cues being embedded within other items in their respective pop-up menus making them inconspicuous. Whereas, the PC cue for IS was placed at the bottom of the list in Experiment 3b. Therefore, with the modified menus, the PC cues (Seat and Boat) were placed at the bottom of the lists for DHC and Trch. To select an item from the pop-up menus involves one moving the cursor (the blue strips in Figure 3.10) upwards from the bottom of the list. By placing the PC cue at the bottom, just like the configuration in the IS menu from the previous experiment, the cue becomes the first item in the list and is not embedded within other items. The PC cue in the modified IS menu was placed at the top of the list. This modification serves as an extra test to assess if the occurrence of PCE in the IS problem is affected by a difference in menu position.

If the obtained differential error reduction effect was due to the relative position of the PC cue in the menus then, with all else being equal, the following predictions are made with respect to the modified pop-up menus:

1) Reliable reduction in PCE occurrences should be observed in the DHC and Trch problem task
2) Reliable reduction in PCE occurrences should still be observed in the IS problem task because despite its modified PC cue position, being at the top of the list, it is not embedded within the other items in the menu.

These modifications made to the pop-up menus serve as an experimental condition (Pop2) and data collected with this condition is compared to the data collected from the previous experiment.

2 Method

2.1 Tasks

The same six problem tasks from Experiment 3b were used.

2.2 Design

A single condition, where participants solve the problem tasks using the modified pop-up menus (Pop2 condition), was run in this experiment. The same counter balancing of testing trials was used as in Experiment 3b.

2.3 Materials

The same computer equipment and paper materials were used as in Experiment 3b.

2.4 Participants

24 participants took part in the current study. They were all UCL undergraduate or postgraduate students. There were 12 males and 12 females, ages ranged from 18 to 30

with a mean of 22.3. Each participant was paid £3 for participating in the experiment; moreover, as an incentive to encourage performance, an extra 50p was paid for every correctly solved problem. Therefore, each participant could get paid a maximum of £6.

2.5 Procedure

The same instructions, training and testing trials were adopted as in Experiment 3b.

2.6 Measures

The same measures of interest were obtained as in Experiment 3b.

3 Results

Analysis was carried out with the data obtained in the previous experiment (conditions Txt and Pop1) and 24 valid cases from condition Pop2 in the current experiment.

3.1 Responses from checklist

Almost all participants (22 out of 24) in the Pop2 condition rated the PC step as "very important" in the checklist. The remaining two rated it as "important". Consistent with responses from Txt and Pop1 in the previous experiment, the PC step was acknowledged as part of the integral solution in the training trial.

3.2 PCE rates in Txt and Pop2

Table 3.9 summarises the number of PCEs and their respective error rates in the two experimental conditions. Almost half of the participants (11 out of 24) in the Pop2 condition committed at least one PCE. Collapsed across problem tasks and conditions (Txt and Pop2), a total of 53 PCEs was generated out of 141 opportunities for the error; yielded an overall PCE rate of 38% between Txt and Pop2.

	Txt	Pop2
No. of PCEs	33	20
No. of PCE opportunities	69	72
PCE rate	47%	28%

Table 3.9: Summary of number of PCEs obtained in the two experimental conditions.

3.3 Difference in PCE occurrences between conditions

A PCE score was computed for each participant in the Pop2 condition. Table 3.10 below shows the respective mean, median and interquartile range of the PCE scores.

	Txt	Pop1	Pop2
Mean	1.43	0.67	0.83
Median	2.00	0.50	0.00
Interquartile range	0.00 – 2.00	0.00 – 1.00	0.00 – 2.00

Table 3.10: The means, medians and interquartile ranges for the three conditions.

The difference in PCE scores between Txt and Pop2 was found to be marginally significant, U = 190.5 (N_1 = 23, N_2 = 24), p = .055. No reliable difference was detected between Pop1 and Pop2, U = 272 (N_1 = 24, N_2 = 24), n.s..

3.4 PCE occurrences in each individual problem task

3.4.1 Comparison between Txt and Pop2

Table 3.11 shows the number of PCEs for all three PC problem tasks (DHC, Trch and IS) in Txt and Pop2.

	Txt			Pop2		
	DHC	Trch	IS	DHC	Trch	IS
No. of PCE	10	9	14	7	6	7
No. of PCE opportunities	23	23	23	24	24	24
Error rate (%)	43.5%	39.1	60.9%	29.2%	25%	29.2%

Table 3.11: Individual problem task's PCE rate in Txt and Pop2.

Planned comparisons using Chi-Square comparisons showed a significant difference between Txt and Pop2 conditions only in the IS problem, $\chi^2 = 4.78$, (1, N = 47), $p = .029$. Descriptively, the difference in PCE rate for the IS problem task was about twice between conditions Txt and Pop2. No significant difference was found in the DHC problem, $\chi^2 = 1.04$ (1, N = 47), n.s., or the Trch problem, $\chi^2 = 1.08$ (1, N = 47), n.s.. The results show a consistent pattern of error reduction with the previous experiment.

3.5 Difference in number of PCEs between the three problem tasks in Pop2

As in the previous experiment, the difference in the number of PCEs are compared across the three problem tasks within condition Pop2. This is to assess if any one of the problem tasks produce a reliably higher PCE rate. Cochran Q test was used to assess the difference in number of PCEs between the three problems. The difference was found to be not significant, Q = .20 (df = 2), n.s.. The result is consistent with the pattern found in Txt and Pop1 in the previous experiment, suggesting that given a particular interface used, the three problems did not differ in terms of the number of PCEs obtained.

4 Discussion

Across the two experimental conditions, Txt and Pop2, the current experiment generated an overall PCE rate of 38% which is comparable to the 35% obtained in the previous experiment. The same pattern of results as the previous experiment was replicated: there was an overall difference in the error rate, albeit marginal, between the two conditions, and the difference was relatively more substantial in IS than the other problem tasks. Although no statistically significant error reduction was detected in the DHC and Trch problems, the trend of the data is consistent as in the previous experiment that fewer PCEs occurred in Pop2.

The findings suggest that moving the PC cue to the bottom of the pop-up menus did not have the anticipated error reduction effect in the DHC and Trch problems. However, as predicted, moving the cue to the top of the menu list in the IS problem did not affect its error reduction effect. The results also suggest no difference in PCE occurrences between the two pop-up menus, namely Pop1 and Pop2. Therefore, the overall findings suggest that the differential PCE reduction effect observed in the previous experiment cannot be due to the relative positions of the PC cue in the menu lists.

This leaves one final speculation, which is the different number of items in the three pop-up menus. The menus in DHC and Trch both have five items but the IS menu only has three. Since the factor of menu size was not controlled for, the speculation implies that the differential PCE reduction between the problem tasks might be related to the differences in the menu size. A possible explanation for the more substantial error reduction in IS could be that the PC cue is made more salient as compared to the PC cues in the relatively larger menus in the other problems. The presence of fewer items in the smaller menu might provide less distraction in relation to the PC cue. As a result, the more salient the static PC cue, the better it is as a reminder of the PC step. This speculation, nevertheless, requires further experimentation before any firm conclusion can be drawn.

However, one would argue that there might be a potential relationship between PCE occurrences and correctness of the solutions. Across the three experimental conditions, more than half of the participants (48 out of 71) had at least one of the three PC problem tasks solved correctly. Table 3.12 shows that there appears to be no substantive difference in terms of the number of PC problem tasks solved correctly between the three conditions.

		Txt	Pop1	Pop2
	DHC	15	18	14
No. correct solution	Trch	5	2	3
	IS	5	6	4
Total		25	26	21
Total no. of PC problem tasks		69	72	72
% Correct		36%	36%	29%

Table 3.12: The no. of PC problem tasks solved correctly in the two experimental conditions.

The completion of a problem task in the current experiments is self-determined by the participants. The participants were explicitly instructed to complete the problem even if they could not arrive at the correct solution. This methodological design is one of practical concern since the problems adopted were not particularly easy when one had to rely solely on one's working memory to solve them. Furthermore, the essence of PCE resides in the fact that one thinks the task has been completed. Table 3.13 shows the proportion of participants in each condition solving the PC problems correctly or not and whether a PCE was made.

Txt

	DHC		Trch		IS	
	PCE	No PCE	PCE	No PCE	PCE	No PCE
Correct solution	9 (39%)	6 (26%)	2 (9%)	3 (13%)	4 (17%)	1 (4%)
Incorrect solution	1 (4%)	7 (30%)	7 (30%)	11 (48%)	10 (43%)	8 (35%)

Pop1

	DHC		Trch		IS	
	PCE	No PCE	PCE	No PCE	PCE	No PCE
Correct solution	3 (13%)	15 (62%)	1 (4%)	3 (13%)	1 (4%)	5 (21%)
Incorrect solution	2 (8%)	4 (17%)	5 (21%)	15 (62%)	4 (17%)	14 (58%)

Pop2

	DHC		Trch		IS	
	PCE	No PCE	PCE	No PCE	PCE	No PCE
Correct solution	3 (13%)	11 (46%)	0 (0%)	3 (13%)	3 (13%)	1 (4%)
Incorrect solution	4 (17%)	6 (25%)	6 (25%)	15 (62%)	6 (25%)	14 (58%)

Table 3.13: Distribution of number of participants solving the PC problems correctly or incorrectly with PCE or no PCE in all three conditions.

While this hypothesis is not one that was originally to be tested in the current study, the relationship between PCE occurrences and correct solutions was explored in each condition for each PC problem task. Fisher's Exact tests were used for each problem task in each condition to assess association between the number of correct solutions and the number of PCEs within each problem task. Table 3.14 summarises the p values for all the comparisons.

	Txt	Pop1	Pop2
DHC	$p = .074$	$p = .568$	$p = .393$
Trch	$p = 1.000$	$p = .446$	$p = .546$
IS	$p = .611$	$p = 1.000$	$p = 1.00$

Table 3.14: p values of Fisher's Exact tests for assessing associations between PCEs and correct solutions for each problem task in three conditions.

There were no significant associations between the number of correct solutions and the number of PCEs observed in any of the problem tasks in all three conditions.

4.1 Limited effectiveness of static visual cues in mitigating PCEs

Using the same game-like procedural task paradigm as Byrne & Bovair (1997), Chung & Byrne (2004) found that effective visual cues in combating PCEs have a couple of properties: first, a dynamic presentation (such as blinking movements) provides an attention-capturing element. Secondly, there is a *just-before* temporal occurrence at the PC step as opposed to it occurring earlier in the task. The temporal characteristic of a cue is explained in terms of the activation-based goal memory (AGM) model (Altmann & Trafton, 2002) suggesting that cues occurring at the beginning of a task tend to be forgotten by the cognitive system when the time comes to execute the PC step. This is because the cue is "masked" by other goals or steps in the task and the activation of the cue would have undergone considerable decay by the time it needs to be recalled. Consequently, the cue would not have enough activation to be retrieved. It was also found that using a static cue which occurred early in the task, but not *just-before* the PC step, did not have an effect in reducing the error.

In contrast to Chung & Byrne's findings, the cues used in the current experiments were static because they have the same appearance as any other items in the menus. In other words, they did not appear any more dynamic relative to the other items. Moreover, these static cues did not have the just-in-time characteristic, i.e. they did not occur just-before the PC step. Nevertheless, data from the current experiment suggest that visual cues which lack the dynamic and just-before temporal properties do have an effect in reducing PCEs. However, the errors still persist and did not decline to 0% occurrence, whereas the use of dynamic and just-before cueing achieved a complete elimination of the error in

Chung & Byrne's study. This supports the notion that static cues without the just-before temporal property are less effective than dynamic cues with the temporal characteristics in combating PCEs.

Contrary to Chung and Byrne's findings, the current findings suggest that a static visual cue might still have an effect, though limited, in mitigating PCEs. There are a number of differences between the current experiments and Chung & Byrne's, and it is possible that some of these differences might contribute to the opposite findings. One prominent difference is the interface of the task environment used. The task in Chung & Byrne's study involves an interface which is visually more dynamic and complex: there are various controls such as buttons, check boxes, indicators and sliders. Interaction with the task environment involves navigating around different parts of the interface. On the other hand, the interface of the pop-up menus is relatively simple, consisting of a couple of menus and text boxes. Interaction with the interface requires only selecting items from the menus. Although the adopted static cues lack the just-in-time characteristic, the relatively static information and simplistic interaction might make the simple static cues salient and, consequently, able to prime the PC step. In contrast, in Chung & Byrne's study, only the dynamic cue was effective because the relatively high complexity of their task environment made the static cue not salient enough to prime the PC step.

The current findings show that a simple static visual cue could reduce PC errors, and the context of the task environment was an important determinant of the effectiveness of the cue.

Overall chapter discussion

Having participants solving mentally demanding problems "in-the-head" imposes high working demands throughout the problem-solving process. The occurrences of PCEs in the current experiments support Byrne & Bovair's (1997) and Mortenson's (2003) general finding that the error rate is sensitive to working memory demands. In the current study, across the three experimental conditions, more than half of the participants (40 out of 71) made at least one PC error. Furthermore, the PCE rate generated in the Txt condition (48%; 33 errors out of 69 opportunities) is comparable to the 50% error rate obtained in Byrne & Bovair's Experiment 2 with a high working memory load condition. This demonstrates that PCEs can be generated reliably under laboratory conditions using a novel experimental paradigm based on problem-solving tasks. The high error rate also confirms the cognitive robustness of the error phenomenon: that it occurs not only in routine procedural tasks but also in problem-solving situations.

The problem-solving tasks used in this study had a number of characteristics that made them particularly effective in provoking PCEs that contrast with the routine proceduralised tasks used by previous studies. First of all, participants were required to solve the problems "in the head", without the use of any form of external representation, which placed a high demand on working memory for keeping track of the problem state. The use of the problem-solving tasks in generating the error is functionally similar to the use of a concurrent secondary task to tax working memory (Byrne & Bovair, 1997), except that the working memory demand is an integral part of the primary problem-solving task, removing the need for an artificial secondary task.

Secondly, Altmann and Trafton (2002) suggest that under normal circumstances, such as in the absence of high working memory load, the relatively low PCE rate in routine procedural tasks (e.g. using an ATM) is due to rote associative learning — in which the temporal order of the steps in a procedural task helps to cue the next step in the sequence. The relatively low occurrence of PCE is explained in terms of the successful cueing from the preceding step to the PC step. The problem-solving tasks adopted in the current study are not routine procedural tasks, so the extent of procedural cueing from one step to the next is minimal. As a consequence, the generation of the error did not require extensive training on the participants' part. Although the participants had training to ensure they had knowledge of the PC step and action, they were not trained in doing the problems *per se*.

The effect of the simple static cue was found to be limited in its effectiveness in that it did not completely eliminate the error across the problem tasks. In relation to Chung & Byrne's (2004) findings, interpretation of the result from the current experiments also suggests that the static visual cues may have a mitigating effect on PCEs depending on the complexity of the task environment. In practice, providing a just-in-time and dynamic cue might not be possible if a system cannot tell where the PC step lies in a particular task. In the current study, completion of the main goal could not be detected by the system, whereas Chung & Byrne's interface could use a just-in-time cue because the system could detect when the main goal had been achieved. The implication is that when it is not possible to provide a cue with the just-in-time and dynamic features, the use of static cues to combat PCE is better than not using any at all.

There are a number of issues that need to be addressed in this study. Firstly, the manipulation between the two different experimental conditions — text boxes and pop-up menus — is not just a difference in the presence of the PC cues but also in their interaction style. This difference in two factors rather than one might be a shortcoming of the experimental manipulation. However, a practical constraint of the manipulation is that a menu-based interaction allows one to select and input an item (in this case, the PC cue) only if the item is actually present in the menu list.

One suggestion for a future experiment is to consider using a text-based interaction style for two conditions — a cue condition and a no-cue condition. In the cue condition, all entities involved in a problem task including the PC cue are to be on screen. In the no-cue condition, only the problem task entities are to be on screen and not the PC cue. This manipulation should control for the difference in PC cue presence while having the same interaction style across conditions. With this modification, it is expected that the same pattern of results to be obtained as the current experiments, namely less PCE occurrences in the cue condition where the PC cue is present on screen.

Secondly, the completion of a problem task in the current experiments is self-determined by the participants. The participants were explicitly instructed to complete the problem even if they could not arrive at the correct solution. This methodological design is one of practical concern since the problems adopted were not particularly easy when one had to rely solely on one's working memory to solve them. Furthermore, it is argued that the essence of PCE resides in the fact that one thinks the task has been completed. There might be a potential relationship between PCE occurrences and correctness of the solutions. While this hypothesis was not tested in the current study, exploration of the data shows there were no statistically significant associations between the number of correct solutions and the number of PCEs in each of the problem tasks across the three conditions.

Finally, one might criticise the approach of analysing the inherent differences among the individual problem tasks as being imprecise in claiming their differential working memory demands. While one could use tools such as Cognitive Complexity Theory (Kieras & Polson, 1985) or building computational cognitive models (e.g. Kotovsky & Kushmerick, 1991) to obtain quantitative estimates of the working memory demands imposed by the different problems, these methods require assumptions about what methods participants use in the problem-solving process. These assumptions cannot be made within the scope of the current study because verbal protocols were not collected. While verbal protocols might be useful in studying problem-solving behaviour, however, this series of experiments used problem-solving tasks as a vehicle to study PCE rather than an investigation into the cognitive processes involved in problem-solving as such.

The current experiments are only the beginning of studying PCEs using a novel problem-solving paradigm. Future research can be directed to adapt and refine the current methodology to address the identified limitations, and some equivocal issues such as the effect of menu size on static visual cues and the relationship between PCEs and correctness of solutions in problem-solving.

The following section is a discussion of the general findings in relation to the existing theoretical approaches to PCE.

Theoretical discussion

The main findings from the series of experiments suggest that PCEs occur not only in routine procedural tasks but also in problem-solving tasks. Moreover, simple static cues were found to have a mitigating effect on the error although complete elimination of the error was not achieved. In this section, these two findings are discussed in terms of the extant PCE approaches in order to examine if these approaches are able to provide some theoretical understanding to the findings.

The AGM approach (Altmann & Trafton, 2002) provides a description of how PCE is avoided rather than how it actually occurs. As discussed earlier in Chapter 2, the AGM model proposes that in routine procedural tasks the avoidance of the error is achieved through associative cueing between task steps. The well-practiced procedures in a routine task act as cues for subsequent procedures, therefore the execution of a PC step is associatively primed by its preceding task step provided it is carried out. However, participants in the problem-solving tasks were not carrying out routine procedures; they did not receive extensive practice in solving the problems. Therefore, associative cueing between task steps in the problem-solving tasks should be minimal if it exists at all. In AGM terms, the susceptibility of PCE in the problem-solving tasks could be attributed to the lack of associative cueing between task steps.

Chung & Byrne (2004) explain their finding of dynamic just-in-time cueing to eliminate PCE in terms of the AGM approach; the just-in-time characteristic of a cue is particularly important since cues appearing early in a task are likely to be masked by subsequent task goals. As a consequence, those cues will not be retrieved in time to prime the execution of the PC step. Current findings that static cues, which lack the just-in-time element, could also reduce PCE might appear contradictory at first glance. However, despite the fact that static cues in the Pop conditions did not occur just before the PC step, they appear whenever the menu was clicked for inputting an item. The cue in the problem-solving tasks then did not appear only at the beginning of the task but throughout the problem-solving process. In terms of the AGM approach, the intermittent appearance of the static cues was able to overcome masking by intermediate task goals and associatively prime the execution of the PC step, though complete elimination of the error was not successful.

The successful generation of PCE using problem-solving tasks supports the general finding that high demands on working memory increase the likelihood of PCE occurrences (Byrne & Bovair, 1997, and Mortenson, 2003). In the current study, participants were required to carry out the problem-solving process without any external aids in the environment, placing considerable demands on the participants' working memory. Byrne & Bovair's CAPS account postulates that as long as the main goal of a task remains active in working memory, activation gets propagated to its subgoals. When the condition of a subgoal is matched and it accumulates, from the main goal, a high enough level of activation, then that subgoal is retrieved and its actions executed. However, completed goals or subgoals decay and get displaced from working memory. Furthermore, the amount of activation that gets distributed to the subgoals varies as a function of working memory load: the higher the load, the less the activation distributed. When the main goal of a task is fulfilled and the system is operating over its working memory capacity, then the main goal will not remain active long enough to propagate enough activation to the PC subgoal. Consequently, the PC subgoal will not get retrieved when its activation level is not high enough to be above threshold. The CAPS account seems able to make sense of the problem-solving situation: the various memory demands, such as having to remember one's state in the problem space and keeping track of the names and number of entities involved, adds to the system's working memory load. When the main goal of completing the problem is fulfilled, it decays and is removed from working memory without propagating enough activation to the subgoal of sending the vehicle back to the other side of the river. However, the CAPS account does not address the influence cues in the external environment have on goal-directed behaviour, which was not the account's original intention. Therefore, it is difficult to articulate the effect static cues have in mitigating PCE in CAPS terms.

The Soar account of PCE is based on a particular property of the cognitive architecture; the asymmetry between the initiation and termination of a subgoal. The initiation of a subgoal occurs when Soar encounters an impasse in a problem space, and the set up of the subgoal leads to other actions which when completed allows resumption to the original main goal. However, the subgoal disappears together with its associated structures, such as a sub-subgoal, once it is completed and terminated. This asymmetry property between goal initiation and termination in Soar does not seem antagonistic to problem-solving situations. Once the main goal of the problem is reached (e.g. all entities get across the river safely) and terminated, its associated subgoal of "sending the vehicle back" is lost and never carried out.

However, PCE does not happen every time and, indeed, participants in the current study did not make the error in every trial. The Soar account further postulates compensatory mechanisms against the error tendency, such as internal cueing through rehearsal strategies reminding one to carry out the PC step, or external cueing from the environment. Furthermore, the compensatory mechanisms are fragile enough that they stop working when influenced by adverse factors, such as high working memory demand. In Soar terms, the compensatory mechanisms might be depleted by the high working memory demand in the current problem-solving task, resulting in successful generation of the error. The mitigating effect on PCE of using static visual cues is consistent with the Soar account's notion of having external cueing as one of the compensatory strategies against the error occurrence.

The supergoal kill-off account bases its theoretical constructs in the construction-integration (C-I) model (Kintsch, 1988). This C-I based account was initially developed to explain behaviour in exploratory learning. Instead of assuming a complete specification of correct task knowledge as in most routine cognitive skills, it is assumed that one's goal hierarchy is incomplete and fragmented. In carrying out a task, the goal structure has a dynamic nature and generation of new subgoals is driven by feedback from the external environment. As far as this assumption of the account goes, it seems applicable to the problem-solving situations because participants would not have had the correct knowledge in solving the problems. However, the difference is that the generation of subgoals during the problem-solving process is most likely to be driven from internal feedback since the participants did not receive any information of the problem states in the task environment; rather the information is in one's working memory. The presence of the static cues in the Pop conditions can be interpreted as information in the external environment which aided the generation of the PC subgoal because it is semantically related to the PC step: the vehicle name relating to the PC step of sending it back.

In explaining PCE, the supergoal kill-off account postulates that as long as the main goal in the hierarchy remains active it supplies activation to its subgoals. When a subgoal is completed its supply of activation is terminated. When two goals are similar, the termination of one might also lead to the termination of the other because they are associatively linked together. The occurrence of PCE is attributed to the high similarity between the completion step in a task and the main task goal; once the completion step is executed, the main task goal also gets deactivated and cuts off the supply of activation to the PC subgoal. Consequently, the PC subgoal never gets executed. Although this descriptive explanation might be stretched to describe the occurrence of PCE in the problem-solving tasks, it does not address the effect of working memory demand imposed by the problem-solving process. Therefore, despite the fact that the supergoal kill-off account appears compatible at first glance with the problem-solving situations and the use of static cues against the error, it does not offer any description to make sense of the effect of working memory demands on PCE occurrences.

In summary, Byrne and Bovair's CAPS account is able to provide a description of the notion of high working memory demand imposed by the problem-solving situation, however, the account is less able to explain the effect of static cues on PCE. This is not surprising given that the original intent of the account was to investigate the effect of working memory on PCE and did not take into account factors from the external environment. The approach offered by the AGM model and the Soar account describe the occurrence of PCE not in terms of working memory demands, but in terms of inherent characteristics of procedural tasks or property of the cognitive architecture respectively. The AGM model explains PCE arising in routine procedural tasks and might not be entirely suitable when applied to problem-solving tasks. However, the description of PCE offered by the AGM model might be stretched to explain how the error occurs in problem-solving tasks due to the lack of associative cueing in non-procedural tasks. The Soar account describes the occurrence of PCE in terms of its architectural property —

asymmetry between subgoal initiation and termination — but it is not clear how the notion of working memory demand might be explained.

The AGM and Soar approaches are able to take into account the effect of static cues since both approaches incorporate the role of the external environment on PCE. Among the two approaches, Soar might be a more suitable candidate for applying to problem-solving situations because the architecture was initially developed to explain problem-solving behaviour. Finally, the supergoal kill-off account takes account of the notion of feedback from the external environment and seems able to describe the mitigating effect of static cues on PCE: the cue generates the PC subgoal through a semantic-action relationship. However, the notion of subgoal similarity, which gives rise to PCE, does not seem compatible with the effect of high working memory demands in problem-solving situations.

The current theoretical discussion has assessed how well the extant theoretical accounts accommodate the findings so far. However, the current series of experiments are not able to distinguish between the different theoretical perspectives. One of the difficulties is that because existing approaches describe PCE using different theoretical constructs, this makes it hard to compare and contrast the different accounts. This is hardly surprising because research on PCE is still in its infancy, although observation and documentation of the error phenomenon can be dated back more than 20 years ago (e.g. Rasmussen, 1982). It is important to examine the existing theoretical accounts closely in order to gain further insights into the nature of the error phenomenon. The insights might then be able to guide further empirical investigation into PCE.

The next chapter is a critical analysis of the four existing approaches to PCE: the supergoal kill-off account, the Soar account, the CAPS model and the AGM model. The objective of this meta-theoretical analysis is to find a common reference point so that the seemingly incommensurable approaches can be compared and contrasted with each other.

Chapter 4

A meta-theoretical analysis of post-completion error

1 Introduction

The empirical studies carried out in the previous chapter extended our knowledge of PCE, that it does not only occur in procedural tasks but also in problem-solving tasks. However, existing theoretical approaches to PCE are scarce and most of the approaches (except the CAPS account) only offer verbal descriptions of the error phenomenon. In order to further pursue PCE empirically, it is necessary to revisit the existing theoretical approaches in more detail to gain further insights into the error phenomenon.

This chapter is a meta-theoretical analysis of PCE. There are several existing theoretical accounts of the error phenomenon, however, they remain rather distant from each other, in that they differ from the each other in terms of their theoretical constructs. For example, although their theoretical positions are similarly grounded in goal management, Polson et al.'s (1992) and Byrne & Bovair's (1997) approaches all have an explicit notion of the activation construct, whereas Young's (1994) Soar account does not rely on such construct. Furthermore, there does not seem to be a consensus among researchers as to what key aspects a model or explanation of PCE should possess. This lack of theoretical consensus can be seen in the different explanations proposed by researchers grounding them in different choices of cognitive architectures such as Soar, ACT-R and CAPS. Although the choice of a particular cognitive architecture is sometimes dependent on the functional behaviour in question, it is, nevertheless, difficult to envisage a coherent theoretical picture of the error phenomenon.

The purpose of the current meta-theoretical analysis is to construct a simple analytical framework, through which to compare and contrast the various existing PCE accounts. However, this is not to claim that the framework should develop into a unified model of PCE: it is constructed and used as a conceptual tool to analyse and evaluate the different proposed theories in a systematic manner. Therefore, as such, the current analysis is best viewed as an attempt to outline what important characteristics an adequate explanation of PCE might need to possess. Moreover, it is also one of the objectives of the current analytical exercise to be able to further empirical investigation through a critical examination of the existing theoretical accounts of PCE.

2 Basis of the analytical framework

The proposed analytical framework consists of a set of criteria and it is argued that a model, or an explanatory description of PCE, should address each of the criteria. The basis of the framework is derived from various theoretical difficulties faced by the different PCE accounts.

Firstly, one of the difficulties in developing a mechanistic model of PCE is to have the model produce the *error rate* that is in accord to what the human subjects produced. Consider even under error-prone conditions, such as having a high working memory demand, subjects in Byrne & Bovair's (1997) Experiment 2 did not make the error on every single trial. As Young (1994) noted "...it's a general problem for *any* deterministic model ... when attempting to account for certain kinds of errors. If a deterministic model makes the error at all, then it will make the error every time. Conversely, if it can do the task right, then it will *never* make the error." (p.10). Young further commented that in order to address this all-or-nothing error behaviour in any explanations one should be aware that, "... errors are typically explained not by a model which actually produces them, but by a model which does the task correctly but about which a secondary argument is made that it is liable or susceptible, in some way, to the error." (p. 10)

The second difficulty, which is related to the first, is addressing the infrequent but persistent nature of the error phenomenon. As described in Chapter 2, Reason (2002) carried out a questionnaire survey asking 95 undergraduates and academic staff to indicate how often they committed various kinds of omission errors when using a photocopier in a naturalistic setting. The participants' responses were based on a 7-point scale ranging from 0 = never to 6 = nearly all the time. It was found that the PCE, leaving the last page of the original, had the highest score among all other omission errors with a mean of 2.18 and a standard deviation of 1.56. Although the PCE was found to be the most prevalent kind of omission error, its frequency score, at least from Reason's study, supports the notion that it does not occur very frequently. It is not unreasonable to suppose that the occurrence of the error in most everyday PC tasks is rather infrequent, since most people are able to carry out PC tasks without making the error most of the time, such as remembering to get change from a ticket machine, retrieving the original after photocopying and collecting one's cash card after a withdrawal on an ATM. However, despite its infrequent occurrence, PCE has a persistent nature in that it happens every now and again. Data from empirical studies, as described in Chapter 2, also lend support to its persistent characteristic that prolonged practice and motivating participants not to make PCE did not achieve complete elimination of the error (Byrne & Davis, in press), PCE was reduced to 0% occurrence only when the task structure was altered such that the PC step has to be carried out before goal completion.

Based on the two inter-related problems posed by theorising about PCE, a set of four criteria proposed by the framework can be formulated as follows:

How the task in question is carried out correctly.

How the error occurs.

How the error is avoided most of the time, i.e., being infrequent.

How the error is sometimes made, i.e., being persistent.

The first criterion: "how the task in question is carried out correctly" is concerned with how a given explanation of PCE deals with the foundational level of the error behaviour; in other words, a description of how the PC task is executed correctly. The second critierion: "how the error occurs" is to do with explaining how the error may arise from the basic foundational level of task performance. The third and fourth criteria: "how the error is avoided most of the time" and "how the error is sometimes made" are concerned with how the infrequent but persistent nature of PCE is explained, and implied in these criteria is also how the issue of an all-or-nothing error rate is addressed.

The four criteria form a set of questions to be asked of each of the extant PCE accounts to examine how each of the criteria is addressed. The nature of this meta-theoretical analysis is qualitative, therefore, there is no strict chronological order in which the criteria should be assessed. In some cases it might be meaningful to assess more than one criterion at the same time.

3 Analysis and evaluation of extant PCE accounts

In the following section, an overview of each of the PCE accounts is presented again, in more detail, and then each of the four criteria from the framework above are used to assess the ways in which each of the theoretical accounts fulfils them (or not).

3.1 Supergoal kill-off account

In suggesting a theory-based evaluation technique for user interface design, Polson, Lewis, Rieman & Wharton (1992) proposed a cognitive model of learning through exploration. At a high level, their model is similar to Norman's theory of action (1988) which describes a cycle of stages in order to carry out some actions in the world and evaluate the effect of the action. Polson et al.'s model extends further and specifies the cognitive mechanisms underlying the control of the different stages in the action cycle. Polson et al.'s model is based on the construction-integration (C-I) model (Kintsch, 1988), which was initially developed to model text comprehension tasks but has been used in areas of HCI as well (e.g. Kitajima & Polson, 1992).

Polson et al.'s C-I based model specifies that the initial goal structure of a user's task can be organised in a hierarchical form similar to ones derived from a GOMS analysis (Card, Moran & Newell, 1983). The goal hierarchy may consist of a top-level goal representing the main task goal, lower-level goals represent the decomposition of the task, and the bottom-level goals represent the executive actions necessary to accomplish the corresponding goals.

Goals and subgoals in the goal hierarchy are represented by propositions and they are connected to each other through a network of associative links which are also represented as propositions. Activations flow from the top-level goal to its subgoals through the associative links and the execution of an action is the result of receiving high enough activation from its associative subgoal. Goals or subgoals become deactivated once they are accomplished and the supply of activation to the subgoals gets terminated when the top-level goal is deactivated, in other words, the subgoals will remain active as long as the top-level goal is active.

Various nodes associating with each task goal exist in the goal structure; such as a "done-it" node which signals the completion status of a goal or subgoal, and an "and-then" node connecting between subgoals specifying the order in which the subgoals are to be executed.

How is the task carried out correctly?

Polson et al.'s model was initially developed to account for behaviour in exploratory learning. Instead of assuming users have a complete specification of the necessary procedures and correct knowledge of the task as in routine cognitive skill, it is assumed that the user's initial goal hierarchy is incomplete and fragmented. Polson et al. suggest that the goal structure has a dynamic nature and is constantly being revised when an action is executed; the generation of new subgoals is driven by feedback from the task environment.

In relation to executing task steps in a procedural task, Polson et al. suggest an "and-then" goal structure in which there are "and-then" nodes interconnecting the various task goals, and these interconnections make up the specific sequence of how the task steps in a procedural task are to be carried out. The "and-then" nodes govern the flow of activations between subgoals; for example, if a first subgoal is connected to a second goal through an "and-then" node, then activation may flow from the "and-then" node to the second subgoal only when the first subgoal is accomplished. Each task goal has an associated "done-it" node, and when the task goal is completed it is deactivated through an inhibitory link from its corresponding "done-it" node.

Although the assumption made about the user's task knowledge is somewhat in contrast to the assumption of correct and complete task knowledge in routine cognitive skills, such as using an ATM, Polson et al. made the assumption because of a different aspect of skill, namely, exploratory learning rather than execution of routine procedural tasks. The assumption does not pose much of a problem in the way they postulate how a procedural task is carried out. In essence, so far, in describing how a procedural task is executed correctly, Polson et al.'s model highlights the important role of feedback from the task environment and the control structure between subgoals through a coordinated dispersion of activations.

How the error occurs?

The occurrence of a PCE within the model is founded upon the notion of *similarity* between the main task goal and the penultimate subgoal, which is usually the subgoal preceding the PC subgoal. For goals that are similar they may be associatively linked to each other through their respective "done-it" nodes; for example, in a photocopying task, the main goal — make copies — may be argued as being similar to the penultimate subgoal — get copies — and their "done-it" nodes, then, associatively link together. When a subgoal is fulfilled and deactivated by its "done-it" node, activation may spread to the "done-it" node of a similar task goal through the associative link deactivating it.

So, in the case of a PC task, such as the photocopying task, the completion of the subgoal of getting copies may cause deactivation of the main goal — make copies — because of their inherent similarity. Once the main goal is deactivated all of its subgoals lose activations and deactivate, because all subgoals remain active by receiving activation from the overall main goal only when it is active. As a result, this premature deactivation of the main goal causes deactivation of the PC subgoal of collecting the original, and the PC step is not executed result in a PCE.

The main problem with the supergoal kill-off account in explaining the occurrence of the error is, as Byrne & Bovair (1997) point out, that the notion of similarity is not operationalised to determine how similar the penultimate subgoal and the main goal must be in order for the error to occur.

How is the error avoided most of the time, i.e., being infrequent? How is the error sometimes made, i.e., being persistent?

The persistent nature of PCE is implied in Polson et al.'s model in that a task with a penultimate subgoal that is similar to its overall goal (the PC goal) will be terminated prematurely every time and never get executed. However, this poses a problem for the account in addressing the infrequent nature of the error: PCE will occur every time if the penultimate subgoal is similar to the main goal but will never occur in a task where the penultimate subgoal is dissimilar to the overall goal. Therefore, this explanation suffers from the all-or-nothing error behaviour which is common to many deterministic models.

3.2 Soar account

Young (1994) used an analytical modelling approach called Programmable User Models (Young, Green & Simon, 1989) to analyse the unselected window error which is similar to a PCE. The Programmable User Models technique is based on the Soar (Laird, Newell & Rosenbloom, 1987) cognitive architecture, and the main idea is to identify the knowledge that a user needs in order to perform the task at hand. Once the knowledge required is made explicit, it can be implemented as rules and run as a predictive model in Soar.

The necessary procedures for carrying out a procedural task are essentially a collection of rules in Soar. These rules are statements, which take the form of IF-THEN conditions, just like many production systems. The rules are organised at different levels representing a hierarchical structure of the main goal, subgoals, sub-subgoals etc., in a task decomposition.

How is the task carried out correctly? How the error occurs?

The way a procedural task is executed in Soar is by treating the task as a problem space, and the model proceeds by applying operators to states in the problem space. When the progress through the problem space is blocked then the model is faced with an impasse. When Soar encounters an impasse it generates a subgoal which when accomplished would enable the model to resume progression in the initial problem space. Once the model resumes to processing of the main goal, the subgoal simply disappears. Young argued that the occurrence of PCE is largely due to this asymmetry between the generation and the termination of a subgoal. The encounter of an impasse causes the generation of a subgoal which initiates some events, however, the termination of a subgoal makes the subgoal itself disappear, and this does not initiate further activities.

A Soar explanation of the photocopying task may have a rule at the main goal level, such as, IF in photocopying room AND have copies in hand THEN go to the scheduled meeting. However, if one is in the photocopying room but with no copies in hand then an impasse occurs and a subgoal of using the photocopier to make copies is then set up. When the subgoal is fulfilled, that is, copies are in hand then control is resumed to the main goal which is to go to the scheduled meeting. Since the subgoal and its associated structures disappear when it is accomplished, the sub-subgoal of collecting the original is lost and is never executed. This subgoal-lost account also implies, implicitly, the importance of the notion of moving on to the next subsequent goal, and indeed, when a PCE occurs in the real world there is usually another main goal to follow. This point may seem rather obvious and trivial; however, it is not addressed properly in most PCE accounts.

Implicit in the way that the Soar account explains how a procedure is executed is also the assumption of a complete specification of the correct knowledge necessary. But the account does not specify, in detail, how the correct order of execution in a procedural task is coordinated. Nevertheless, the explanation of PCE occurrence is based on an inherent property of the architecture. So how does the task ever get executed correctly given the model is faced with the problem of only producing the error behaviour?

How is the error avoided most of the time, i.e., being infrequent? How is the error sometimes made, i.e., being persistent?

Although the Soar account offered by Young was not implemented as a running model, it was suggested that correct task performance needs explanation that goes outside the model itself. It was suggested that the use of an internal cue, such as self-reminding, or an external cue, such as some cues in the task environment, may ensure the rule fires at the right time to carry out the PC subgoal. The internal and external cues may be used discretely or in combination at the same time; these strategies, which lie beyond the model, help account for how the task is carried out correctly without making PCE.

While the use of internal and external cues help explain the capability of correct task performance of the model, the account still needs to overcome the problem of not generating the error behaviour. It was further specified that the internal and external cueing strategies are sufficiently fragile in that they are susceptible to a wide range of adverse conditions such as time pressure, high working memory demands and general distractions. Therefore, while the architectural property of Soar lays down the foundation addressing the persistent nature of PCE, the fragility of the compensatory strategies for correct task performance addresses how the infrequent nature of the error phenomenon may arise within the model.

3.3 Limited working memory capacity account

In accounting for their empirical findings that PCEs are more likely to occur in low-capacity subjects under high working memory demands, Byrne & Bovair (1997) offered an explanation implemented as a running computational model based on Collaborative Activation-based Production Systems (CAPS) (Just & Carpenter, 1992). The basic conceptualisation of Byrne & Bovair's PCE account is similar to Polson et al.'s supergoal kill-off approach, in that goals are structured in a GOMS-like hierarchy. Subgoals remain active by deriving activation from the top-level goal as long as it, too, is active. The occurrence of the error is attributed to the loss of activation supply to the PC subgoal when the top-level goal is accomplished. However, instead of postulating goal similarity as the cause of the error, Byrne and Bovair operationalise the construct of working memory capacity in CAPS and are able to specify the condition under which the error is more likely to occur, namely, high working memory load.

There are several memory components responsible for the basic mechanics of CAPS: working memory, production memory and long-term memory. The working memory system stores dynamic information about the task at hand, and each memory element is associated with an activation value which has to be above a threshold level in order to be active in working memory for processing. If a memory element has an activation level below the threshold it is effectively dropped from working memory. The flow of activation between working memory elements is controlled by production rules in the production memory system. Productions are matched against working memory elements and only active elements may satisfy the conditional side of a production. When productions are matched they can fire repeatedly to propagate activation to a memory element which has below-threshold activation, or they can fire to request activation for maintaining active elements in working memory. Productions in CAPS may fire in parallel when more than one production is matched at any one time. There are declarative memory structures in the long-term memory system, and these memory structures are not subject to activation constraints such as decay.

The essence of CAPS is a limited capacity of the working memory system. The limited capacity is instantiated as the total maximum amount of activation available to working memory. In general terms, an individual who has a capacity of 52 units is said to have a higher working memory capacity than an individual who has 24 units of activation in capacity. Constraint of this limited capacity in working memory is manifested through trade-offs between storage demand and processing demand. Activation is required for both the maintenance of active elements in working memory (i.e. storage) and the propagation of activation to other inactive memory elements (i.e. processing). Suppose a CAPS model is working at its capacity limit and more activation is required for operation, then the amount of activation for both storage and processing will be cut back proportionally by a back-scaling mechanism to keep the total amount of activation available to the limit. For example, if the total capacity available in working memory is 40 units of activation and 30 units are required for maintenance of old working memory elements, and 20 units are requested for propagating activation to inactive memory items, the required amount of activation is 10 units more than the capacity limit. The back-scaling mechanism then cuts back both the maintenance and propagation request by 40/50 (i.e. 4/5), so 24 units are allocated for maintenance and 16 units for propagation. This back-scaling mechanism has two consequences in terms of the behaviour of a model. First, less activation than requested is obtained for maintenance of old elements in working memory. This means some of the working memory elements will lose their activation (i.e. decay) and get displaced if their activation levels are below threshold. This results in forgetting due to decay of memory items. Second, the amount of activation received for propagation to inactive memory items is also less than the requested amount. Consequently, in order to boost the activation of the inactive items up to threshold the corresponding productions have to increase the number of firing cycles. This leads to a slowdown in processing due to increased computation. This constraint imposed by storage and processing demands only applies when demands on a CAPS model's working memory exceed its capacity limit.

How is the task carried out correctly?

The limited capacity model addresses performance of routine procedural skills by assuming the presence of a complete specification of correct procedural knowledge of a task. The procedural knowledge in Byrne & Bovair's CAPS model was derived from a GOMS analysis, and was translated into a set of productions in the production memory system.

Similar to the supergoal kill-off approach, one of the assumptions in Byrne and Bovair's CAPS model is that the supply of activation to a task's subgoals depends on the active status of the top-level goal. As long as a task's top-level main goal remains active in working memory, subgoals can receive activation from the firing of a production responsible for goal maintenance. On each processing cycle the goal maintenance production fires, propagating activation to a subgoal in order to keep it active in working memory until it is accomplished. This goal maintenance mechanism enables the CAPS model to execute a specified task procedure in the correct order.

Under circumstances when working memory is not taxed, in other words, the CAPS model is working within its capacity limit, the task will be carried out without making a PCE. If the model's working memory capacity is not stressed there is no scaling back of activation for the use of maintenance and propagation. Therefore, when a task's top-level goal is accomplished it is kept active in working memory without any decay to its activation value. This enables the goal-maintenance production to fire and propagate activation to the PC subgoal until its activation level reaches threshold, and subsequently, it becomes active in working memory and carried out.

However, as Altmann & Trafton (2002) point out, since completed goals do not get dropped out of working memory immediately in CAPS, correct task performance of Byrne & Bovair's model depends on a certain level of working memory demands. The decay mechanism in CAPS is not time-dependent but load-dependent. The decay of a memory item's activation value only occurs when another memory item requests activation. When the model is processing with insufficient working memory load, accomplished goals will not get displaced to prevent them from interfering with unaccomplished goals. This is because goals in CAPS may still be matched when completed as long as it is still active in working memory (this is in contrast to Polson et al.'s supergoal kill-off account where completed goals are deactivated immediately). Therefore, a CAPS model has to work with a minimum demand on working memory so that completed goals decay and get displaced from working memory to ensure the correct sequence of actions to be executed. However, people seem to be able to perform tasks, in general, correctly without a minimum demand on working memory.

How the error occurs?

Byrne and Bovair's CAPS model specifies that PCE occurs when working memory demands exceed the model's capacity limit. The occurrence of PCE in CAPS is due to the back-scaling mechanism that takes place when the model's working memory is stressed. The back-scaling mechanism results in a cut back of activation for maintenance of old working memory items, which may include an accomplished top-level goal, and propagation of activation to the to-be-carried-out PC subgoal. If the accomplished top-level goal remains active in working memory, the goal maintenance production propagating activation to the PC subgoal will have to increase its number of firings in order to boost the subgoal's activation to threshold. This is to compensate for the cut back in the amount of activation being propagated resulting in a slowdown of processing. However, if the cut back of activation for maintenance involves the top-level goal being displaced from working memory before the PC subgoal accumulates enough activation to be above threshold, then the goal maintenance production stops firing and the PC subgoal stops receiving activation. Consequently, the PC subgoal does not get carried out because it is not active in working memory, resulting in a PCE.

How is the error avoided most of the time, i.e., being infrequent? How is the error sometimes made, i.e., being persistent?

Byrne & Bovair's account of PCE addresses, partially, the infrequent and persistent elements of the error phenomenon by being able to make predictions about *when* the error is going to occur: in this case, under high working memory demands. However, this is only a partial success because the model still faces the problem of all-or-nothing error generation. The behaviour produced by the model follows that "with a high enough working memory capacity, the error does not occur, and with low capacity the error is always made." (p. 41). Although, in Byrne & Bovair's study, significantly more PCEs were found with low-capacity individuals performing an experimental task with extra working memory load, the error did not occur in every single trial.

3.4 Activation-based goal memory account

Altmann & Trafton (2002) proposed an activation-based goal memory (AGM) model arguing that goal-directed cognitive behaviour can be explained, parsimoniously, by treating goals just as ordinary memory elements rather than some special memory structures (e.g. Miller, Galanter & Pribram, 1960; Goschke & Kuhl, 1993). The AGM model is developed to model processes in goal selection in the Tower of Hanoi task and is based on the ACT-R cognitive architecture (Anderson & Lebiere, 1998) but rather than using the entirety of the architecture, ACT-R's goal stack mechanism was "turned off" and the model was implemented using the basic memory mechanisms of the architecture.

Using the construct of activation, the AGM model suggests that, just like other memory elements in the cognitive system, goals have associated activation levels and cognition is directed by the most active goal retrieved at any given time. This central idea is elaborated through three main components in the model: the interference level, the strengthening process and the priming process.

In addition to a retrieval threshold already present in ACT-R, which specifies that a given memory item may be retrieved successfully only if its activation level is above the retrieval threshold, the AGM model incorporates an interference level, which is a threshold value imposed by competition among memory items. The interference level is the mean activation value of the most active distractor item, and the target item must have an activation value higher than the interference level in order to get retrieved successfully. The amount of activation associated with a memory item is subject to decay, and this decay process is time-based and gradual. The gradual nature of the decay process means that recently retrieved items may still have relatively high activation values during the early stage of the process, and interfere with the retrieval of a target item.

In order to overcome interference from distractor items, the AGM model postulates that a target item must undergo a strengthening process to become the most active item. The strengthening process is a process in which a new goal is encoded; it is during this process that the activation of a new memory item gets initialised. In more descriptive terms, the strengthening of an item is akin to "focusing attention" on it. During strengthening, an item's activation gets accumulated rapidly until it is above the interference level, then it gets retrieved and directs behaviour. However, the system cannot spend too long strengthening a target item because the more active it is now the more it will interfere with retrievals of later items. The AGM model specifies that the time the system spends strengthening a memory item is determined by the retrieval threshold (taken to be the activation value when the item is initialised) from below and the interference level from above. The strengthening process has to be long enough so that the target item gets active enough above the interference level, but not so long that its activation persists and interferes with retrievals of new goals at a later stage. However, there is noise in the activation values of memory items, so even when a target item has activation above the interference level, this does not guarantee retrieval. Therefore, there is possibility that, due to noise, the activation level of the target item goes slightly below the interference level and does not get retrieved on a given cycle of processing.

Once a goal takes over and directs behaviour, it begins to decay gradually. When the system shifts to a new goal, the old goal continues its decay process and becomes less and less active. If the cognitive system needs to refocus attention to (or resume) an old goal then this old goal needs to undergo a priming process to become active again. On the one hand, the priming process is similar to the strengthening process in that it accumulates activation for a memory item. On the other hand, the priming process specifies that for an old item to be reliably retrieved later, it has to be primed associatively from a retrieval cue to boost activation above the interference level. An associative link between a retrieval cue and an old goal must be established in the first place in order for the retrieval cue to effectively prime the old goal when it comes to resumption. In other words, the associative relationship between a retrieval cue and an old goal must be learned before goal suspension, so that the occurrence of the retrieval cue can be associated with the suspended goal upon resumption. A retrieval cue can be internal, residing in the cognitive system; a procedural task step can act as a cue for its following subsequent task step. Therefore, task steps in a learned procedural task can be viewed as a sequence of associative links priming the execution of their subsequent steps. A retrieval cue can also be external, residing in the environment; for example, a loud beeping signal in the ATM when it returns the cash card can prime the action of collecting the card back provided the relationship between the cue (the beep) and the action (collection of card) is learned.

How is the task carried out correctly?

According to the AGM model, Altmann & Trafton suggest that although the decay process is functional because it prevents the perseverance of a memory item's activation from interfering with later retrievals, there is also a cost with the decay process. Unlike Byrne & Bovair's limited capacity account, the decay process in the AGM model is based on time rather than the load on working memory; therefore, the longer in time a goal is suspended, the more it decays, and consequently the longer it takes to be retrieved successfully upon resumption.

In relation to PCE, Altmann & Trafton suggest that because the PC subgoal, by definition, is the final step in a procedural PC task, the PC subgoal would have been suspended for the longest and so decayed the most. As a consequence, the default tendency for the cognitive system is to make the error. However, it is argued that the fact people seem to be able to perform PC tasks, in general, without making the error most of the time suggests the priming process, in the AGM model, provides "deliberate cognitive operations" (p. 64) to overcome the default tendency to err at a PC step. The priming process ensures the execution of procedural subgoals in the correct order through associative priming between consecutive subgoals. The reliable execution of the PC subgoal is because its preceding subgoal serves as a retrieval cue when it is time to carry out the PC action. This notion of associative priming between subgoals has also been applied to model airline pilots' visual scanning behaviour in a procedural task (Schoppek, 2000). Schoppek shows that even low-level cognitive processing, such as visual scanning within a procedural task setting, can be modelled as a set of associative links between actions: execution of an action acts as a cue spreading activation to its subsequent action through the associative linkage.

The AGM model provides a description of correct procedural task performance and its time-based decay process addresses one of the limitations in Byrne & Bovair's CAPS model, which requires a constant minimum level of working memory load to ensure correct task performance. However, the assumption of the default tendency for the cognitive system to make PCE due to the PC subgoal being suspended longest is problematic. By the same logic, subgoals that are near the end of a procedural task with a considerable number of task subgoals are also likely to be omitted, because these subgoals would have been suspended long enough to undergo substantial decay, and this argument should not be specific to the PC subgoal. Moreover, the proposal of the PC subgoal being suspended requires the assumption that all subgoals of a task were somehow set up in working memory in the first place. While the notion of goal suspension might be applicable to tasks such as Tower of Hanoi, where suspension of subgoals are necessary in some steps in order to proceed in solving the problem, it is not clear whether it is necessary or how all the subgoals in a procedural task need to be set up and suspended in working memory at the beginning of the task.

A further shortcoming of the AGM model as an account of PCE is that although it makes the important distinction of internal and external cueing in task execution, it does not make explicit the importance of a false completion signal in a PC task. The presence of a false completion signal in a PC task is a significant contribution to the occurrence of PCE, as discussed later in the current chapter. Although some of the assumptions made by the AGM model might be problematic as an account of PCE, it provides a clearer and more detailed theoretical foundation than the CAPS account in terms of the execution of procedural task steps.

How the error occurs?

The PCE account offered by Altmann & Trafton explains how the error is avoided more adequately than explaining how the error occurs. Altmann & Trafton applied their account tentatively to explain Byrne & Bovair's general result that PCEs occur under high working memory load. It is suggested that because the decay process in the AGM model is time-based, when there is a high working memory load more items are to be encoded by the system per unit time. This leads to increased interference in memory retrievals. However, the explanation is problematic in two ways; firstly, it can easily be interpreted as an explanation for errors in general occurring in procedural task execution rather than as an explanation specific to PCE. This is because, unlike the CAPS account, the AGM explanation does not take into account the notion of task completion, which is the essence of PCE that the error occurs after a task is completed. Secondly, even if the proposed explanation was to apply specifically to PCE occurrence, it implies that instead of retrieving the PC subgoal the cognitive system is likely to retrieve goals that were retrieved recently because they would have suffered less decay. However, it is not unreasonable to describe that, in everyday life, most PCE situations usually involve one omitting the PC step and pursuing the next goal rather than repeating an action one had already carried out.

The AGM model is able to provide some theoretical and empirical leverage in generating PCE under conditions such as interruption (see Chapter 6), though as an account of PCE it has its limitations.

How is the error avoided most of the time, i.e., being infrequent? How is the error sometimes made, i.e., being persistent?

Altmann & Trafton explain that the natural tendency of making PCE is avoided by means of some "deliberate cognitive operations", more specifically, the associative cueing between procedural task steps. Furthermore, the AGM account is similar to the Soar account of PCE, though starting from quite different mechanisms, that the persistent nature of the error phenomenon is explained by the models' default tendency of making the error. However, as mentioned earlier, the AGM model as an account of PCE describes better how the error is avoided than how it is generated.

4 Discussion

Having described and assessed each of the existing PCE accounts against the criteria generated from the analytical framework, Table 4.1 is a summary of how each of the accounts addresses the specified analytical criteria. In the Table, a *Yes* represents that a criterion has been addressed by a particular account and a *No* represents the criterion has not been addressed by the account. There are a number of short comments in brackets in the table; they highlight the extent to which the criterion is addressed by a particular explanation. Based on the summary table, the following section compares and contrasts the different accounts in relation to the specified criteria.

	Correct task performance	Occurrence of PCE	Infrequent but persistent nature
Supergoal kill-off	*Yes*	*Yes*	*Yes (but only partially)*
Soar	*Yes (but not in detail)*	*Yes*	*Yes*
Limited working memory capacity	*Yes (but problematic assumption)*	*Yes*	*Yes (but only partially)*
AGM	*Yes*	*No*	*Yes (but only partially)*

Table 4.1: How each of the extant PCE accounts addresses the criteria specified in the analytical framework

4.1 Compare and contrast of the extant PCE accounts

4.1.1 Correct task performance

Common to all the extant accounts of PCE is that they all address how the task at hand is executed correctly; all four accounts cast their explanations in terms of goal management and they agree with the assumption of a hierarchical structure of goal organisation in procedural PC tasks. Furthermore, apart from the Soar account, all other approaches adopt the theoretical construct of activation in explaining the dynamics within the goal structure. The difference in the Soar account is mostly an inherent characteristic of the architecture, which is most suitable for modelling problem-solving behaviour. However, when applied as an explanation for procedural task execution, the Soar account does not offer more details than treating a procedural task as a problem space, and the model traverses this space by applying operators to different states.

Most of the accounts, apart from the supergoal kill-off account, have a consensus that PCE is not due to a lack of correct knowledge; in other words, the presence of correct knowledge of the procedural task at hand is one of the assumptions made. Although the underlying theory of the supergoal kill-off account was not initially set out to assume the presence of complete and correct task knowledge, this is not thought to pose a particular problem to the account in terms of its subsequent explanation of PCE occurrence. However, a complete account of PCE should also be able to describe and explain behaviour of the task in which the error occurs. Given most PCE situations that have been documented (e.g. Reason, 1990; 2002; Rasmussen, 1982) and empirically investigated (e.g. Byrne & Bovair, 1997) involve procedural task execution, the supergoal kill-off approach might be problematic in describing the underlying task execution behaviour.

Byrne & Bovair's limited capacity account did not explain in more detail *how* the task procedures were coordinated to give rise to the correct sequence, apart from assuming the presence of the correct knowledge. Furthermore, correct task performance in Byrne & Bovair's model depends on the assumption of a working memory load minimum, which is problematic when people seem to be able to achieve error-free performance under no working memory demands.

The AGM approach provides a more parsimonious and detailed description of procedural task execution than other existing accounts of PCE. For example, the Soar approach does not give detail about task execution; the proposal of "done-it" and "and-then" nodes to coordinate action sequences in the supergoal kill-off approach is rather arbitrary; and the assumption of minimal working memory load during task execution in the limited capacity account is problematic as discussed earlier. In contrast, the proposal of procedural task execution offered by the AGM approach is theoretically grounded in existing memory structures and mechanisms in a cognitive architecture, namely ACT-R.

4.1.2 Occurrence of PCE

A commonality between the Soar and the limited capacity account is that they both suggest that the occurrence of PCE is due to certain cognitively adverse conditions. The limited capacity account is able to operationalise and specify one precise condition under which PCE is likely to occur, namely, high working memory load. However, the Soar account only provides general conditions such as distractions or shift of attention.

The supergoal kill-off account explains the generation of PCE in terms of similarity between subgoals. However, the notion of goal similarity lacks specificity and is, therefore, difficult to operationalise and specify when the error is going to occur.

When applying the AGM model to explain PCE, Altmann & Trafton do not account, in detail, for how the error may arise from correct task performance. The explanation offered by the account describes more how PCE is *not* made rather than how it is generated.

4.1.3 Infrequent but persistent nature

In addressing the infrequent but persistent nature of PCE, the problem is overcoming an all-or-nothing error rate faced by any deterministic model of human error. This problem is addressed in all of the existing PCE accounts, though rather differently to each other. Although the supergoal kill-off account is capable of an error-generating mechanism within the model's scope, the mechanism suffers from the all-or-nothing error problem; when the goals are similar the error happens every time, but if the goals are dissimilar then the error never occurs. Similarly, the limited capacity account is capable of predicting when the error is likely to occur; however, the CAPS model is deterministic in that it still suffers from all-or-nothing error behaviour.

The Soar account is the first account that attempts to address the infrequent but persistent nature of PCE. The persistent nature of the error is explained by Young's analysis on the tendency of the Soar cognitive architecture naturally making the error. The infrequent nature of the error phenomenon is explained in terms of the compensatory strategies to overcome the error. Similarly, though starting from quite different theoretical premises, the AGM model addresses the persistent nature suggesting a natural inclination in the cognitive system to make the error due to decay of suspended goals. The system is capable of avoiding the error by some "deliberate cognitive operations" like priming between procedural steps. A commonality between the Soar approach and the AGM approach is a three-layer form of argument that addresses the all-or-nothing error behaviour: first, a natural tendency for the model to make the error; secondly, the presence of compensatory strategies or mechanisms to avoid the error; and thirdly, those strategies/mechanisms might be fragile enough that they get disrupted under cognitively adverse conditions. However, both approaches are rather vague in specifying the nature of the conditions that provoke the error. Of the two approaches, the AGM model provides a more detailed theoretical description of procedural task execution and this can be extended to make empirical predictions about the effect of interruption on PCE in a procedural task paradigm. The investigation of the effect of interruption is dealt with in the next chapter.

4.1.4 False completion signal and a follow-on task

Further to the number of criteria initially set out by the analytical framework, the analysis also makes explicit two important attributes of PCE; *false completion signal* and a *follow-on task*. The notion of a false completion signal in a PC task is not new: it was pointed out by Reason (2002) in a study analysing what characteristics in a photocopying task make the PC step more susceptible to omission relative to other omission errors. However, this notion is not made explicit in most of the extant cognitive accounts of PCE, apart from the CAPS model. Although the supergoal kill-off account and the AGM model have explicit notions of the roles external cues and feedback from the environment play in task execution, no reference was made to the presence of false completion signals in PC tasks. In the AGM model, Altmann & Trafton suggest that cues used for associatively priming the next task step can be internal (to the cognitive system) or external (in the environment). Similarly, most PC tasks have false completion signals in the environment, for example, the output of photocopies, the withdrawal of cash, the dispensing of a train ticket etc., and this false completion signal might act as a *miscue* priming the execution of the next task.

While it is argued that the presence of a false completion signal might be a source of cueing to pursue the next task, a closely related notion, namely a follow-on task, has not been addressed explicitly in most of the current PCE accounts. Most PCE situations in everyday life involve one moving on to a follow-on task, such as going to a meeting when you have the photocopies, going to a restaurant when you have got the cash etc., and the idea is that the follow-on task might act as a source of "attractor" as against the PC step from a previous task. The notion of a follow-on task is implied in the Soar account suggesting the resumption of a main goal when a subgoal is completed and vanished. However, all other proposed explanations treat PC tasks as "stand-alone" tasks without taking into account the significance of goals/tasks following the PC task.

4.2 What can we learn from the current analysis?

Although it is out of the scope of the current analytical exercise to specify a complete account of PCE, the following outcomes from the meta-theoretical analysis suggest possible starting points for further theoretical development of the error phenomenon. Firstly, except the Soar approach, all existing accounts do not explicitly address the all-or-nothing problem in explaining the occurrences of PCE. In order to overcome this problem, the current analysis suggests that an adequate account of PCE might require a three-layer explanation to describe: (1) how the task in question is carried out, (2) how PCE arise from the task performance behaviour, and (3) how the infrequent but persistent nature of the error is manifested.

Secondly, the analysis suggests that the AGM approach provides the most detailed and theoretically driven description of procedural task execution when compared to the other approaches. Therefore, the AGM model might serve as a good theoretical starting point to investigate PCE within the context of a procedural task.

Thirdly, in attempting to explain the occurrence of PCE, all current approaches give accounts in terms of goal management in the cognitive system. However, the current analysis suggests that attributes of PC tasks, such as a false completion signal and a follow-on task, might also be important factors provoking the error.

Fourthly, to address the persistent and infrequent nature of PCE the current analysis suggests incorporating a level of fragility into the compensatory mechanism(s) that avoid the error in the first place. The level of fragility should be sensitive to the condition under which the error is observed. The exact nature of fragility requires further research; however, mechanisms within a probabilistic framework might serve as a starting point.

5 Concluding remarks

A meta-theoretical analysis on extant accounts of PCE has been carried out by identifying a number of criteria under an analytical framework. The analytical framework enables a systematic assessment of the existing accounts identifying their commonalities as well as their differences. Furthermore, by comparing and contrasting under the framework it has been possible to bring the seemingly disparate accounts under a coherent structure.

In addition to the critical assessment exercise offered by the current analysis, it is possible to outline how an adequate account of PCE might require a three-layer explanation to describe: first, how the task in question is carried out; secondly, how PCEs arise from the task performance behaviour; and, thirdly, how the infrequent but persistent nature is manifested. Furthermore, the analysis also suggests that the notion of a false completion signal and a follow-on task might be significant attributes present in the error phenomenon.

Despite the fact that it is not possible to develop a unified model of PCE in terms of detailed cognitive mechanisms with the current analysis, this meta-theoretical analysis allows one to think about the error phenomenon within a coherent theoretical structure. By postulating high-level descriptions, it serves as a good stepping-stone to further future theoretical development. In order to achieve a low-level detailed account of PCE, it is necessary to complement the current analysis with the hard work of developing running computational cognitive models. Developing a computational cognitive model is not a trivial task and it is beyond the scope of the current thesis; however, future research should take the ideas from the current analysis into the modelling process.

All in all, the current theoretical explanations of PCE are still in their infancy, and this is also true of empirical investigation of the error as indicated by the small number of studies carried out. This might be because there are insufficient empirical studies to support substantial theoretical development, or it might be the other way round: because there are no substantial theoretical accounts guiding empirical research of the error phenomenon. Nonetheless, one of the contributions of the current analysis suggests that the AGM model is a good theoretical candidate for guiding empirical studies of PCE in a procedural task context. Consequently, the subsequent chapters constitute an empirical examination of PCE, and an attempt to apply the AGM model to help generate predictions of the effect of interruption on PCE in a procedural task context. Chapter 5, is a review of research on interruption and this is followed by experimental studies of the effect of interruption on PCE in Chapter 6.

Chapter 5

Literature review of interruption research

1 Introduction

Interruption in the work place is common and can occur quite frequently, especially in an open-plan office environment, such as colleagues dropping by, the telephone ringing, other people talking on the phone, etc. As our modern working environment becomes more reliant on computers and there is a need for efficient communications, interruptions become more prevalent. Just to give a few common examples from office desktop applications: notification of newly arrived emails, instant messaging, and even pop-up advertisements while surfing the web!

The effect of interruption has been studied from a range of different perspectives from occupational psychology to HCI. The general theme of interruption research addresses disruption to performance of an ongoing main task caused by some interrupting activity. The disruptive effects of interruption have been studied in simulated office environments with professional office workers (Zijlstra, Roe, Leonora & Krediet, 1999); although interruptions generally have a negative impact on main task performance, some positive effects, such as, increased speed of task execution have also been observed. The effect of interruptions have also been studied in computerised clerical tasks (Burmistrov & Leonova, 1996), instant messaging (Cutrell, Czerwinski & Horvitz, 2000; Czerwinski, Cutrell & Horvitz, 2000), data management in telephone operators (Eyrolle & Cellier, 2000), and professional information workers (Czerwinski, Horvitz & Wilhite, 2004). Common to these example studies is that they address the effect of interruption in work environments or using work-related computer applications. The general finding that emerges from these studies of interruption is that it affects main task performance in a negative way, such as slower task performance, increased effort and negative effect on emotions. These negative effects might mean decreased efficiency and productivity in an office setting; however, in a safety-critical environment the effect of interruption might have far more severe consequences.

In the nuclear power industry, it has been suggested that most of the incidents involved the operators' attention being distracted away from the primary task (Bainbridge, 1984). In the aviation domain, Dismukes, Young and Sumwalt (1998) carried out a review of the National Transportation Safety Board (NTSB) reports and concluded "...nearly half of these accidents [attributed to human error] involved lapses of attention associated with interruptions, distractions, or preoccupation with one task to the exclusion of another task." Dismukes et al. also pointed out that pilots often get interrupted, such as by messages from air traffic control, during flight-deck checklist operations. Degani and Wiener (1990) noted that there were three aviation accidents, in a period of fifteen months, involving omission of items in a checklist and all three accidents involved crashes not long after takeoff. Latorella (1999) carried out an empirical study of the effect of interruption on pilot's flight-deck performance. The study revealed that interruptions to flight-deck procedural tasks significantly impaired performance by increasing error occurrences; it was found that pilots are 53% more likely to commit an error in a procedure when being interrupted. As Latorella noted, although some of the errors committed due to interruption might not be operationally significant, some might have life-threatening consequences.

Clearly, a lot of errors during task execution can be related to interruption. Reason (1990) noted that omission errors are especially prone to interruption, suggesting that this is because the interrupting activity might be "counted in" as part of the interrupted task sequence, or the interruption causes the operator to "lose his place" upon resuming the interrupted task. Given that PCE belongs to the bigger family of omission errors, its relation to interruptions is worth investigating in order to gain a better understanding of the error phenomenon.

The organisation of the following literature review begins with some early basic psychological research that started investigating interruption as a topic; this serves as an introduction to the history of the research topic. This is then followed by recent psychological studies related to interruption, namely task switching and prospective memory research. To a certain extent, task switching research illustrates its relevance to interruption research in terms of a general cognitive cost when switching between tasks, however, the experimental paradigm in task switching does not have a clear distinction between a primary task and an interrupting task. On the other hand, the experimental paradigm used in prospective memory research provides a clearer distinction between primary and interrupting tasks. Results from several recent studies investigating interruption are then presented and, as described later, these results, for example on task similarity and complexity, are often conflicting because of a lack of a common theoretical tool to study interruption. This is then followed by introducing Altmann and Trafton's (2002) AGM model again because it has been used as a theoretical framework by a number of recent studies to examine different dimensions of interruption such as rehearsal strategy, cue availability and the time it takes to resume an interrupted task. Finally, findings of interruption duration and position are reviewed and the AGM model is discussed to make predictions about the effects of interruption duration and position on PCE.

2 Early basic psychological research on interruption

The first series of psychological studies related to interruption research date back to the work by Zeigarnik (1927). Zeigarnik's research focus was on the effect of interruption on memory performance; interruption was used as a manipulation to study the memory preservation of uncompleted tasks that have been interrupted and the memory of completed tasks. Zeigarnik gave her participants a series of tasks to perform ranging from manually oriented ones, such as paper folding, to mentally oriented, for instance multiplication tasks. The participants were instructed to finish the tasks as rapidly and accurately as possible. Among these tasks, half of them were interrupted before completion by the experimenter presenting the next task to be carried out, and half of them were not interrupted. Zeigarnik found that when the participants were asked to recall what the tasks were, the ones that were interrupted before completion were better recalled than the ones that were completed. This effect of better memory retention for uncompleted activities than completed activities is known as the "Zeigarnik effect" (Van Bergen, 1968).

Ovisankina (1928) carried out subsequent studies related to Zeigarnik's but with a different behavioural focus. Ovisankina was interested in whether task resumption occurs after an interruption, and if it does under what conditions it occurs. The participants in Ovisankina's experiment were given a series of tasks to carry out, and some of the tasks were interrupted "deliberately" by the experimenter presenting the next task to be performed; while some other tasks were interrupted "accidentally" by the experimenter, such as, asking the participant for help after dropping a box of pins. The participant was then left alone for some time after an interruption during which the experimenter observed his/her behaviour. Ovisankina found that all accidentally interrupted tasks were resumed and almost 80% of deliberately interrupted tasks were resumed within 20 seconds. It was also observed that resumption occurred most frequently when a task was interrupted at the beginning rather than the middle or towards the end of the task.

Early psychological accounts of the results from Zeigarnik's and Ovisankina's studies proposed a tension system (reviewed in Van Bergen, 1968) in which when an intention is formulated, tension is set up and the system tries to resolve that tension by fulfilling the intention. When a goal is formulated to complete a task, there is tension in the system and the tension is discharged when the task is completed. An interrupted task means tension release is blocked and the system is still under tension, and this may have an effect on task resumption and memory. The resumption of an interrupted task is viewed as an attempt of the system to release the tension by completing an uncompleted task. Alternative theories of task interruptions from early psychological principles, such as Gestalt psychology and psychoanalysis, have also been proposed. However, a review of these early theoretical accounts is beyond the scope of the current thesis.

Although more recent research on interruption, as discussed later in the chapter, is concerned more with the disruptive effect it might have on some primary task performance, the early studies have shown a number of features of interruption that have inspired subsequent research on the phenomenon: first, interruption affects behaviour in terms of memory processes; secondly, the affected behaviour also manifests itself in subsequent behaviour *after* the interruption; and thirdly, an interruption may manifest itself in different ways and affect behaviour differently.

3 Recent psychological research related to interruption

3.1 Task switching

In the domain of attention and performance, psychological research into cognitive control has been investigating the psychological processes underlying task switching: a situation where a person is alternating between different tasks. The general findings in the task switching literature, as summarised by Monsell (2003), suggest that there is a cost associated with task switch, this cost is measured immediately after switching to a different task and manifests itself in terms of increased reaction time (RT) and decreased accuracy of task performance. Furthermore, although this switch cost can be reduced when participants receive a brief period to "prepare" for a task switch, it does not eliminate the switch cost completely; there is always a cognitive overhead associated with switching task.

Rogers and Monsell (1995) developed an experimental paradigm known as the alternating-runs paradigm to investigate task switching. The paradigm involves two simple cognitive tasks, such as, a letter-discrimination task and a digit-discrimination task. The two tasks are presented on a computer screen and switches between the two tasks take place in a predictable manner; in other words, participants know when the task switch is occurring. Task cues are usually present to indicate when the task is going to switch. The paradigm allows switch cost to be estimated by comparing RTs in switch trials to RTs in non-switch trials. The general findings show that there were significant increases in RT and errors between task switch, and the switch cost (in terms of time) decreased when participants were given up to a 0.6-second preparatory time before task switch. However, the switch cost still persisted even when the preparatory period was increased to 1.2 seconds. The beneficial effect of a preparatory period to reduce switch cost was also replicated by Gamble and May (2002) using the alternating-runs paradigm.

Rogers and Monsell explain switch cost in terms of a penalty imposed by a mental "gear change" to enable reconfiguration of cognitive processes for the new task. It is suggested that there are two control processes involved in the reconfiguration: endogenous and exogenous process. The endogenous process is an internally-driven, top-down process which is adopted freely at will before presentation of the next stimuli in a sequence. The exogenous process, on the other hand, is a data-driven, bottom-up process which is triggered by the presence of stimuli. Rogers and Monsell suggest the persistent switch cost that cannot be eliminated by a preparatory period is likely to be due to the exogenous process, which cannot be carried out until the presentation of stimuli occurs.

Instead of a reconfiguration process through a mental "gear change", Wylie and Allport (2000) suggest switch cost can be explained by negative priming or proactive interference from a previous task. It is argued that the emphasis of the observed switch cost is on a preceding task rather than an upcoming task and that the effect does not need to come from the immediate preceding stimuli only.

Gamble and May (2003) examined switch cost associated in switching between task sequences rather than a single task unit. Modification was made to the alternating-runs paradigm using graphical stimuli-discrimination tasks. Gamble and May manipulated switching between two task sequences each consisting of four subtasks, and the primary interest of this manipulation was to investigate where the switch cost would occur in a task sequence. It was found that switch cost occurred at the first subtask of a task sequence, suggesting that the observed switch cost was associated with an entire task set rather than individual task steps. This result seems to lend support to the interference account of task switching because the reconfiguration account would predict that switch cost would occur between subtasks in a task sequence.

There is ongoing debate between the different theoretical accounts and experimental manipulations (e.g. Altmann, 2003, 2004; Monsell, 2003) which is not within the scope of this thesis. Nevertheless, the phenomenon of task switching is related to interruption at a high-level because interruption involves switching from an interrupted task to an interruption, and vice versa upon resumption of the primary task. However, there are some differences between task switching research and interruption research. Firstly, the magnitude of observation in task switching is on a much smaller scale, usually within hundreds of milliseconds, whereas the behavioural scale (if measured in RT) in interruption research is usually on the seconds level. Secondly, the paradigm used in task switching does not make a distinction of primary and secondary task: the paradigm consists of rapid presentation of stimuli alternating between two tasks. In interruption research, the effect of an interruption is usually measured on some aspect of performance on a primary task. Finally, experiments in task switching usually involve tasks with very simple and abstract stimuli, but interruption experiments usually adopt tasks with a higher level of complexity such as game-like task environments (e.g. Gillie & Broadbent, 1989; Trafton, Altmann, Brock & Mintz, 2003). Despite the differences between the two related research topics, namely task switching and interruption, the main message to take away from the task switching literature is that there is a cognitive cost associated with performing more than one task serially.

3.2 Delayed-execute prospective memory

The effect of interruption has also been investigated in the field of prospective memory (PM) research. Recent studies have examined the effect of interruption using an experimental paradigm called delayed-execute PM. The idea of a delayed-execute PM task is to simulate a situation where an intention is retrieved but cannot be carried out immediately. For example, a real world situation might be that you are to pass on a message to your colleague when you see him/her in the office building. However, when you see him/her, he/she is engaged in a phone conversation so you have to wait until the phone conversation is over before you can pass on the message to him/her, so there is a delay between the retrieval and the execution of the intention.

Einstein, McDaniel, Williford, Pagan and Dismukes (2003) investigated the effect of interruption and divided attention on delayed-execute PM performance. The experimental paradigm involved participants performing an ongoing primary task, and when they encountered a PM cue (when the computer screen turns red) the participant was required to retrieve an intention (pressing the slash key on a keyboard) which is to be carried out at an appropriate time (when the ongoing primary task is completed and changed to a different task). The period between the onset of the PM cue and the appropriate time to carry out the intention is the delay, and during this delay participants were required to carry on with the primary task. Einstein et al. manipulated the length of the delay (5 seconds vs 15 seconds vs 40 seconds) and had a condition where there was an interrupting task during the 40-second delay. The interruption was a pattern-comparison task lasting 15 seconds, so the total delay length was still 40 seconds for this condition. A further manipulation was assigned such that some participants were asked to perform a digit-monitoring task as well as the ongoing primary task during the delay, whereas some participants did not have the digit-monitoring task. The results showed that PM performance was worse when the participants had to carry out an attention-dividing task (digit-monitoring) during the delay. Inclusion of an interrupting activity during the delay did have a significant detrimental effect on PM performance relative to no interruption during the delay. Moreover, the effect of interruption was more pronounced when the participants' attention was not divided, suggesting that the effect of divided attention overshadowed the effect of interruption during the delay. However, there was no effect of delay length on PM performance.

The disruptive effect of interruption on delayed-execute PM performance was replicated by McDaniel, Einstein, Graham and Rall (2004) using the same experimental paradigm. McDaniel et al. further manipulated the duration of the interruption during the delay period. The two interruption durations tested were 10 seconds and 20 seconds. The findings suggest that when there was an interruption during the delay, PM performance was significantly worsened; however, there was no differential effect between the two interruption durations. McDaniel et al. concluded that interruption duration (at least for the ones tested) did not have a differential effect on PM performance; however, there is a methodological shortcoming in their investigation. The two different interruption durations (10 seconds and 20 seconds) were embedded within a 40-second delay so that the total delay length was unchanged. However, the design of the manipulation is such that there was a delay period *after* the interruption: 20 seconds for the 10-second interruption duration and 10 seconds for the 20-second interruption duration. The presence of these post-interruption delay periods might allow participants to develop compensatory strategies to overcome any differential effects imposed by the different interruption durations. This suggests that a period as brief as 10 seconds after an interruption might be long enough to overcome the disruptive effect of the interruption. A refined methodology could reduce the post-interruption delay lengths to a minimum (such as 1 or 2 seconds) and also make them the same length for the two interruption duration groups.

The investigation of the disruption effect of interruption within the delayed-execute PM paradigm is similar to most studies in interruption research that there is a clear distinction between a primary task and an interrupting task. The main general finding from the PM paradigm is that PM performance can be affected negatively by the presence of interruption. Furthermore, studies in the PM literature did not measure the disruption effect immediately after an interruption and this suggests that an interruption might have a persistent residual effect, at least in the case of PM, some time after its occurrence.

4 Some recent studies on task similarity and complexity

Studies that treat interruption as a research topic in its own right have started investigating different dimensions of the phenomenon. The following sections provide a review of the main dimensions that have been examined.

One of the dimensions of interruption that has been investigated is the nature of the interrupting task and its relation to the interrupted task. In experimental psychology research, Gillie and Broadbent (1989) have examined several characteristics of interruption including task similarity and complexity. The primary task adopted in their study is a computer-based task involving participants memorising a list of items to be collected from various specific locations in a game-like environment. The lists of items to be collected vary in number (5 items or 7 items) in order to manipulate memory load imposed by the task. In examining the effect of task similarity, Gillie and Broadbent found that an interruption involving encoding and recall of stimuli (memory-intensive, hence similar to the primary task) is more disruptive than an interruption task of simple numerical addition (processing-intensive but dissimilar to the primary task). The effect of task similarity was also found in other studies (Edwards and Gronlund, 1998) suggesting more disruption results from an interruption when it shares more information with the primary task.

However, Gillie and Broadbent suggest that similarity is not the only factor determining the disruptive effect of the interruption task. When a dissimilar, yet complex, mental arithmetic interruption (involving decoding alphabets into numbers) was used, performance on the primary task was significantly impaired. Regarding task complexity in terms of primary task memory load, there was no performance difference between the 5-item and 7-item conditions. This led Gillie and Broadbent to conclude that both similarity and complexity of an interrupting task can have detrimental effects on primary task performance. Furthermore, it was suggested that memory load imposed by the primary task does not seem to make an interruption more disruptive.

From more applied research in HCI, Bailey, Konstan and Carlis (2000) have also investigated the effect of task similarity. Bailey et al. used a range of web-based tasks involving counting, addition, image comprehension, reading comprehension, registration-type information entry, and word matching as primary tasks to examine disruptiveness from interruptions involving reading comprehension and decision-making. However, in contrast to one of Gillie and Broadbent's findings, the effect of similarity between the interruption and the primary task was found to be not significant. The conflicting results obtained might be attributed to the relatively less controlled methodology in Bailey et al.'s study compared to Gillie and Broadbent's. The notion of similarity was ill-defined in Bailey et al.'s study in that both interrupting tasks have substantial overlap in terms of the kind of cognitive processing involved, because both tasks consist of elements of comprehension and decision-making. This makes detecting differential effects between the two interruptions difficult. On the other hand, a result consistent with one of Gillie and Broadbent's results was obtained: that high memory load at the point of interruption does not make an interruption more disruptive to the primary task. However, there is a shortcoming in Bailey et al.'s study that, unlike Gillie and Broadbent's, memory load imposed by the primary task was not formally operationalised in terms of task demand such as number of operations that need to be carried out. Therefore, face value of the results suggest consistency but caution needs to be taken in order to draw firm conclusions from it.

In terms of complexity of the primary task, a recent study, however, has obtained conflicting results. Speier, Vessey and Valacich (2003) suggest that interruptions disrupt complex primary task while facilitating simple task performance. The behavioural focus of Speier et al.'s study is accuracy and pace in decision-making tasks. The simple primary tasks used in the study involve information acquisition, simple addition or subtraction calculations, and identifying trends in data. The complex primary tasks involve order ranking of options after assessing a fairly large number of relevant parameters and carrying out some calculations; in some tasks participants are also required to memorise outcomes from their previous performances. The interrupting tasks adopted are simple information acquisition tasks. The general results show that interruption has a disruptive effect on complex primary task performance: decisions made are slower and less accurate relative to performance in simple primary tasks. The results also show that interruptions occurring in simple primary tasks facilitate performance in terms of faster decision times relative to non-interrupted simple tasks. Speier et al. conclude that while interruptions can facilitate performance on simple tasks, complex task performance can be hindered by interruptions. The facilitation of simple task performance by interruptions is attributed to increased levels of stress narrowing the decision-maker's focus of attention. Ratwani, Trafton and Myers (2006) also found speed-up of actions with the presence of interruptions in a simple task and suggest that the observed benefit might be due to improved perceptual processing in interrupted situations.

The problem with the above studies producing conflicting or incompatible results about certain dimensions of interruption, such as task similarity and complexity, is not only because of differences in their experimental methodologies; the lack of a common theoretical framework providing commensurable guidance also contributes to the incoherent results obtained in the studies. Although there are theoretical attempts to characterise and describe different dimensions of interruptions for investigations (e.g. Latorella, 1999; Walji, Brixey, Johnson-Throop & Zhang, 2004; Speier et al., 2003), these frameworks have their origins in the applied HCI domain and do not provide enough specificity to make predictions about the cognitive processes involved in examining the phenomenon. Although general memory theories in cognitive psychology have been applied to HCI research studying interruption (e.g. Oulasvirta & Saariluoma, 2004), the theories and the applications of them were not able to make predictions about the specific processes involved in resuming an interrupted task.

Recent studies, however, have been able to utilise a common theoretical framework to guide the design of an experimental paradigm examining interruption, and produce harmonious results that can be explained in terms of similar theoretical constructs. One such cognitive theoretical framework that is particularly useful in studying interruption is the AGM model (Altmann and Trafton, 2002). The model has been described earlier in Chapters 2 and 4, but the following section re-introduces the most pertinent aspects of the model to interruption.

5 A common theoretical framework in interruption research

In contrast to the traditional view of goal memory as having a "special status" in the cognitive system (Anderson and Liebere, 1998), the AGM model treats goals the same way as other memory elements, having activation values and being subject to decay. The goal with the highest level of activation is retrieved by the cognitive system and directs behaviour. Other active memory elements that have been recently retrieved or associatively linked to the context in use may compete in the retrieval process. This competition forms an interference level (or a "mental clutter") which gives rise to potentially retrieving the wrong goal. When an interruption occurs, a goal from a primary task is suspended in order for the cognitive system to direct attention to deal with the interrupting task. During the interruption, the activation level of the suspended goal is subject to a time-based decay process if it is no longer in use. Successful retrieval of the suspended goal upon task resumption depends on its activation value being above the interference level and this, in turn, depends on whether some kind of rehearsal activity is possible for the suspended goal during or before the interrupting activity. If the suspended goal is rehearsed throughout the interrupting period, this might counter substantial decay to its activation and still be the most active goal upon primary task resumption.

The AGM model provides several recent studies with theoretical tools to further investigate and explain the seemingly contradicting empirical findings in interruption research. In examining task complexity in interruptions, Hodgetts and Jones (2005) found evidence consistent with Gillie and Broadbent's findings that complex interruptions have more disruptive effects on primary task performance than simple interrupting activities. Hodgetts and Jones adopted a computer-based problem-solving task called Tower of London (ToL) as the primary task. The ToL task is similar to the Tower of Hanoi task (see Simon (1975) for more details on the task) but the discs involved are all of the same size in order to do away with the disc-size constraint, which imposes a limit that a disc cannot be placed on top of a smaller disc. The discs are of different colours so the challenge of the ToL task is to move the discs around so that the configuration of the disc colours matches a desired goal-state configuration. The ToL task is essentially a simpler version of the Tower of Hanoi puzzle. Using the ToL task, Hodgetts and Jones (2006a, 2006b) carried out a series of experiments investigating different aspects of interruption effect; therefore, multiple citations of the two papers will appear in the various sections that follow according to which aspect of interruption is being discussed.

In one of their experiments (Hodgetts & Jones, 2005), participants were required to plan and memorise their solutions and solve the ToL task from memory; the discs on the computer screen were concealed. Two interrupting tasks were used: the simple interruption involved a straightforward checklist of the participant's mood state, and the more complex interruption involved a verbal reasoning task. The findings suggest that participants took significantly longer to resume the ToL task after being interrupted by the complex verbal reasoning task than after the simple-checklist interruption. The same result was replicated even when the participants were not required to solve the ToL task from memory: that is, the discs on the computer screen were not concealed (Hodgetts & Jones, 2006a). Hodgetts and Jones explain their finding in terms of the AGM model: that a complex interrupting task involves more subgoals than a simple interruption, and the larger number of subgoals imposes greater retroactive interference to the cognitive system upon task resumption. Monk (2004) investigated the effect of frequent versus infrequent interruptions and suggests that one might expect more detrimental performance from frequent interruptions due to increased retroactive interference. However, it was speculated that if the interrupting task is not cognitively demanding then enough goal rehearsal might be possible during the interruption to mitigate the adverse effect of frequent interruptions. The AGM model is, thus, able to explain the effects of interrupting tasks of different complexity in a coherent fashion in terms of interference and goal rehearsal opportunity.

6 Rehearsal strategy in interruption

There are only a handful of studies that touch upon the nature of goal rehearsal and its beneficial effect in relation to interruption. Articulatory rehearsal of a piece of memory may help maintain it in the cognitive system (Baddeley, 1990); however, the beneficial effect of verbal rehearsal was not found in PM performance (Einstein et al., 2003). Einstein et al. instructed their participants to verbally rehearse the PM task during a delay before the intention could be carried out in the delayed-execute paradigm. During the delay, participants were to carry on with the primary task and perform a concurrent attention-demanding task. However, no significant improvement on PM performance was found even though the participants received explicit instruction to rehearse the required action. This led Einstein et al. to conclude that this verbalisation strategy might not alleviate the disruptive effect of interruption while participants' attention was divided. However, caution should be taken with Einstein et al.'s conclusion: in their study participants who did not receive the rehearsal instructional also indicated that they had rehearsed the required action in a post-experiment questionnaire. Although the participants who received the rehearsal instruction indicated more rehearsal ratings, the fact that all participants had employed rehearsal strategies might have already benefited their observed performance. Therefore, the explicit instruction given to some of the participants might not have any additional benefit. This does not lend support to the claim that verbal rehearsal is ineffective in mitigating the disruption effect of interruption.

Recent interruption research seems to suggest that rehearsal of a suspended goal may take the form of non-verbal activity. Gillie and Broadbent (1989) argue that a mental arithmetic task with two digits is demanding enough to suppress verbal rehearsal and, yet, results from one of their experiments suggest that performance on the primary task is not affected by the interruption. This led Gillie and Broadbent to conclude that information about the primary task may be retained in another form of temporary memory storage rather than a verbal form. In doing the ToL task, Hodgetts and Jones (2006b) asked their participants informally if they verbally reminded themselves about the move they were supposed to carry out after an interruption. Most participants indicated that they tried to remember the spatial relations between the pegs rather than verbally rehearsing the move sequence to be executed. In support of this claim, Trafton, Altmann, Brock and Mintz (2003) analysed their participants' verbal protocols and found that the information to be retained over the interrupting period was always in reference to cues in the external task environment. This emphasises the important role of environmental cues in resuming an interrupted task.

7 The effect of cue availability on interruption

The AGM model also makes predictions about the effect of cue availability on interruption. The priming constraint in the model specifies that in order for a cue to be effective in priming the to-be-resumed goal, associative links between that goal and the cue have to be formed before the interruption (goal suspension) and that the cue must be present upon goal resumption. Moreover, the cue does not have to be in the external environment but can also be a mental cue in the internal cognitive system. Altmann and Trafton (2004) carried out an experiment manipulating the presence of cues before the occurrence of interruption. Participants were provided brief periods (2, 4, 6 or 8 seconds) before the interruption and during these periods the computer screen "froze" so that no cue was present during this "frozen" period; in the no-cue condition, the computer screen was blank. It was found that cue availability was only effective in mitigating disruption from the interruption in the 6-second and 8-second "frozen" period. Altmann and Trafton suggest that it might take the cognitive system 6 to 8 seconds to establish associations between cognitive representations and the external environment. However, one of the weaknesses of Altmann and Trafton's results is that they compared performances across experiments; this suggests subsequent research should include a full factorial design to avoid possible confounds from changes in subject population.

Hodgetts and Jones (2006b) examined the effect of cueing before, during and after interruption using the ToL task paradigm. In one experiment, participants were provided with a number of interruption conditions: one interruption covered only part of the computer screen so the primary task environment was still visible during the interruption; one full-screen interruption occluded the primary task environment completely; and one full-screen interruption had a 2-second "frozen" period during which the primary task environment was visible but no actions could be performed, and during the interruption the primary task was completely invisible. It was found that the time it took to resume the primary task in the full-screen interruption with a 2-second "frozen" period was quicker than the full-screen-interruption-only condition, but not significantly quicker than the partial-screen-cover condition. Contrary to Altmann and Trafton's finding, Hodgetts and Jones suggest that even as brief as a 2-second cue can be effective in resuming the primary task. Moreover, Hodgetts and Jones's experiment provides stronger evidence as they carried out their manipulation in a single experiment. In a further experiment, Hodgett and Jones manipulated cue availability after interruption by changing the colour of the discs. The results showed that the change of disc colours after interruption had a detrimental effect resulting in slower performance in resuming the primary task. Hodgetts and Jones explain the result within the AGM framework, suggesting that the impaired resumption performance is due to disruption of associative links between the to-be-resumed goal and the environmental cues upon task resumption.

8 Detecting the disruptive effect of an interruption using more sensitive measures

Grounding analysis in low-level memory processes, the AGM model also has implications in terms of designing experiments to investigate interruptions. Altmann and Trafton (2002) suggest that the low-level view of goal memory offered by the AGM model allows one to dissect the temporal structure of an interruption under fine-grain detail. In essence, the occurrence of an interrupting event can be viewed as three temporally related components: first, the occurrence of an alert which does not necessarily demand immediate engagement with the interrupting activity, such as the ringing of a telephone; secondly, is the interruption proper in which engagement is required with the interrupting activity itself, such as the telephone conversation — the period between the alert and the interruption proper is termed *interruption lag*; thirdly, is the resumption of the interrupted task — the period between interruption proper and the re-engagement of the primary task is termed *resumption lag*. This temporal characterisation has direct implications for measuring and designing the dependent and independent variables in examining the effect of interruption.

Using resumption lag to measure the disruptiveness associated with an interruption is suggested to be a more sensitive measurement (Altmann and Trafton, 2002) than global measurements such as task completion time (e.g. Gillie and Broadbent, 1989 and Bailey et al., 2000). Measuring resumption times allows one to assess the disruptiveness *immediately* after an interruption and prevents participants using compensatory strategies to maintain primary task performance, which might not be reflected in overall task completion time. When manipulating the length of interruption lag as an independent variable, Trafton et al. (2003) found that participants given an 8-second interruption lag significantly reduced their resumption times when compared to participants who were required to engage with the interruption immediately. The primary task adopted by

Trafton et al. is a resource-allocation task and the interrupting task is a forced-choice decision-making task. Consistent results are produced across other task domains, such as the problem-solving ToL task. Results suggest that interruption lag as short as 3 seconds followed by a verbal reasoning interruption (Hodgetts and Jones, 2003), or 2 seconds followed by a mood-checklist interruption (Hodgetts and Jones, 2006b) are still effective in reliably reducing the resumption times. The AGM model attributes the benefits of an interruption lag to the opportunity for the cognitive system to associatively encode the to-be-suspended goal with the cue necessary in resuming that goal.

9 The effect of interruption duration

One of the dimensions of interruption that illustrates the particular relevance of the AGM model is the duration of an interrupting event. Previous studies, such as Gillie and Broadbent (1989), investigated interruption duration of 30 seconds versus 165 seconds and found no differential effect in terms of disruptiveness. The primary task used is a computer-based memory game (as mentioned earlier) and the interrupting task is a double-digit mental arithmetic task with its duration varied. Gillie and Broadbent conclude that duration alone does not seem to determine the disruptiveness of an interruption. Bailey et al. (2000) also found that interruptions of different lengths do not contribute to the disruptiveness of an interruption. However, one of the main weaknesses of these two studies is the way they measured the disruptive effect of an interruption: global measures such as overall task completion times were taken as an indication of disruptiveness. As mentioned in the previous paragraph, the use of global measures cannot be an accurate indication of disruptiveness because participants could engage in various compensatory strategies after an interruption. This highlights one of the benefits offered by the AGM model, which is the use of resumption times measuring how long it takes a participant to re-engage with a primary task immediately after an interruption.

The effect of interruption duration is one of the main focuses of the current study and the AGM model is particularly useful in making predictions about its effect. The gradual time-based decay process in the AGM model specifies that the activation level of a goal memory is subject to decay through time. If a primary task goal is not in use (or suspended) during an interrupting event its activation level decays more the longer the interruption duration. In order for a suspended goal to undergo decay the interruption has to be cognitively demanding enough to prevent goal rehearsal. Therefore, the AGM model predicts two consequences with interruption duration: first, when an interruption duration is long enough for a suspended goal to decay below an activation threshold level, as specified by the interference level, it takes longer to retrieve the suspended goal successfully than when an interruption duration is so short that the suspended goal has not undergone substantial decay below the interference level. This is because the cognitive system takes time to boost up the activation level of a suspended goal through the strengthening and priming processes, as proposed by the AGM model. Secondly, an alternative consequence is that an incorrect competing goal is retrieved, leading to an error, when the interruption duration lasts long enough for substantial decay to occur. Monk, Boehm-Davis and Trafton (2004) compared resumption times after interruption durations of 0.25 second, 1 second and 5 seconds. The results show that the 5-second interruption condition yields a significantly longer resumption time than the two shorter interruption conditions. Although resumption times for the 0.25-second and the 1-second conditions are not significantly different, both are reliably longer than a control condition with no interruption. Monk et al. suggest that not only do different interruption durations have differential disruptive effect, even interruptions as brief as 0.25 seconds can result in a cost in resumption time.

The effect of interruption duration is also obtained in a recent study (Hodgetts and Jones, 2006a). Hodgetts and Jones adopted the ToL task as the primary task, and a mood status checklist as the interrupting task. The durations tested were 6 seconds versus 18 seconds; only one checklist was to be completed in the 6-second condition and three checklists in the 18-second condition. A control condition with no interruption was also incorporated. In order to ensure their participants were retrieving an old goal after an interruption and not re-planning the puzzle solution, the participants were instructed to spend 8 seconds to plan and memorise the solution before solving the task. It was found that the 18-second condition resulted in significantly longer resumption times than the 6-second condition, and resumption times from these two conditions were reliably longer than those in the control condition.

Recent studies examining the effect of interruption duration are able to obtain results that are consistent with the prediction made by the AGM model. Moreover, using resumption time as a sensitive behavioural measure, these studies suggest the opposite to the claims made by earlier studies about the null effect of interruption duration.

10 The effect of interruption position

Another dimension of interruption of interest to the current investigation is interruption position, that is, where in the primary task the interrupting event occurs. Adamczyk and Bailey (2004) examined the effect of interruption at different moments within task execution. A range of computer-based tasks including editing, media usage and searching were used as primary tasks. A reading comprehension of news extracts task was used as the interrupting task. An event perception technique (Zacks & Tversky, 2001) was used to elicit participants' mental models of the structure of a task. The technique involved showing some sample tasks to the participants and having them decide where major and minor breakpoints are in the tasks. Based on the task models elicited from the participants, predictions of "best" (least disruptive) and "worst" (most disruptive) moments in the primary tasks were made; for example, in the editing task the "best" interruptive moment was upon completion of an edit like correcting a spelling mistake, whereas the "worst" interruptive moment was during an edit like typing or selecting some text. The results show that interrupting the participants at the "worst" moments during primary task execution can lead to changes in emotional state, such as increased annoyance, and changes in social attitude such as losing respect for the primary task. The implication of this study is that interruption position can have an effect on primary task execution.

Monk, Boehm-Davis and Trafton (2002), and Monk & Trafton (2004) obtained results that suggest differential disruptive effects of interruption position. In their study, participants were asked to program a simulated VCR as the primary task. The interrupting task was a perceptual-motor task, which involved tracking an on-screen target with a mouse. The interruption occurs every 5 seconds and the authors categorised the interruption occurrences into three positions for analysis: start-subtask (at the beginning of a new subtask), mid/end-subtask (in the middle or near the end of a subtask), and scrolling (subtask involving repetitive scrolling operations). The dependent measure of disruptiveness was resumption time after interruption. The findings show that mid/end-task interruptions result in significantly longer resumption times than start-subtask and scrolling interruptions. Monk et al. explain the findings in terms of the AGM model suggesting that the encoding demand in the middle (or towards the end) of a subtask is higher than at the beginning of a subtask or in a repetitive operation. The higher encoding demand results in longer resumption times because more information needs to be rehearsed about the goal state and its associative links to the next task step. This finding is consistent with Adamczyk and Bailey's that, across the range of primary tasks (editing, media usage and searching), the "worst" moments for interruptions to occur are during execution of a task. The "best" moments to minimise disruption effect from an interruption are after completion of a task.

The effect of interruption position was also found in the ToL problem-solving task (Hodgetts and Jones, 2005). Participants were required to first plan and memorise the solution of the problem then execute the solution moves from memory. The optimal solution to the problem was four moves. Two interruption positions were tested: the first move and the third move. Disruptiveness of an interruption was measured by resumption time after the interruption. The results suggest that interruption occurring before the first move yields longer resumption time than interruption occurring before the third move. Hodgetts and Jones suggest that at the beginning of the problem there are more competing goals active in the cognitive system than towards the end of the problem. The more the competing goals, the more the interference during goal retrieval, resulting in longer resumption time. This empirical finding, however, is the opposite of Monk et al.'s finding which suggests interruptions occurring at the beginning of a subtask result in less encoding demand than in the middle of a subtask. Although there appears to be inconsistency between the two findings, explanations offered by Monk et al. and Hodgetts & Jones suggest consistency in terms of the AGM model. The seemingly different empirical findings between the two studies are more likely to be attributable to differences between the primary tasks used. In solving the ToL task, participants were required to memorise the solution moves at the beginning of a trial. This suggests a high load of working memory is involved before the first move can be executed. On the other hand, the VCR paradigm used in Monk et al.'s study is, essentially, a display-based task which a participant's working memory load could be "offloaded" to the task environment. While encoding demand might be higher in the middle of a subtask because, for example, more information on the display has to be differentiated, there might be less working memory demand at the start of a VCR subtask than at the beginning of the ToL task. Whatever the ultimate explanation is, the important message from these findings is the significant effect of interruption position on primary task performance. Moreover, the AGM model can accommodate these findings in terms of differential encoding demands and interference imposed by different points in a task.

The effect of interruption position and the perceptual processes involved upon task resumption have also been investigated by a recent study using an eye movement tracking technique (Ratwani and Trafton, 2006). Instead of using a task with a hierarchical goal structure (Monk et al., 2004), Ratwani and Trafton used a rather linear task with a flat goal structure as the primary task. The task involved categorising numbers on a spreadsheet application. The interrupting task is a single-digit addition task and interruption positions were either early (occurring early in the task) or late (occurring nearer the end of the task). Results on resumption times suggest that there were no significant differences between early and late interruptions. This finding is contrary to earlier results on resumption times in relation to the effect of interruption position, but the null effect is most likely due to the nature of the primary task with a flat goal structure. Ratwani and Trafton suggest that instead of retrieving specific subgoal information, analysis on eye movement data shows that the participants were able to use spatial memory to resume the primary task after an interruption. Moreover, the spatial heuristic used for task resumption was more imprecise for later interruptions than early interruptions. This study shows that different interruption position in a linear task may also have differential disruptive effect on spatial memory upon task resumption.

11 Summary and the relations between interruptions and PCE

The literature on interruptions shows that studies ranging from applied HCI research to cognitive psychology research have examined different dimensions of interruption such as task similarity, complexity, the nature of primary task rehearsal, cue availability upon resumption, interruption position and duration. Some of the findings and explanations appear to be inconsistent or even sometimes contradictory between different studies. This is mainly because of the lack of a common theoretical framework in early interruption research to guide the design of experiments, and to explain findings in terms of similar cognitive constructs. However, the recent development of the AGM model (Altmann & Trafton, 2002) has helped rectify the situation and has inspired a number of controlled experimental studies examining different dimensions of interruption. The AGM model is not only able to provide theoretical tools to explain various findings in a coherent fashion, but also to offer a fine-grained view of interruptions in terms of low-level memory processes. This fine-grained dissection of interruption has enabled the use of resumption time to measure the disruptiveness of an interruption, and arguably it is a more sensitive measure than other global measures, such as overall task completion time that has been used in early studies (e.g. Gillie & Broadbent, 1989).

As discussed in the literature review earlier, recent studies have used resumption time as the dependent measure in examining various aspects of interruption. The successful adoption of this dependent measure is best illustrated in the investigation of interruption duration; studies that used global measures such as the total time spent on primary tasks conclude that there was no differential disruptive effect of different interruption durations (Gillie & Broadbent, 1989; Bailey et al., 2000). However, findings from recent studies suggest that interruptions of different durations result in differential disruption when measured locally using resumption time (Hodgetts & Jones, 2006a; Monk & Trafton, 2004).

Although research on interruptions is becoming more coherent and mature in terms of providing explanations of various findings and the use of experimental methodologies, there is only one study (Mortenson, 2003) that has explored the effect of interruption on a specific kind of error, namely PCE. Mortenson implemented a brief interruption message that the participants were sometimes required to read and dismiss just before a PC step in the robbery game task. It was found that the interrupting message did not increase the likelihood of PCE. However, a methodological shortcoming in Mortenson's experiment might contribute to the null effect of interruption on PCE: the interrupting message itself was found to be easy to ignore by participants with little cognitive demands placed on them. As described earlier, previous studies (e.g. Hodgetts & Jones, 2005; 2006a and

Monk, 2004) suggest that simple interrupting tasks are not as disruptive as complex interruptions in terms of primary task performance.

Given that various statistics and reports have indicated that interruption can be one of the main factors resulting in human error, and sometimes vital accidents, it seems a logical and fruitful avenue to examine the effect of interruption on PCE, in more depth relative to Mortenson's study.

The AGM model is a particularly useful theoretical framework in terms of examining the effect of interruptions on PCE for two reasons. First, its suitability to interruption research has been shown by recent studies which adopt the AGM model and are able to make predictions about various dimensions of interruption effect. Secondly, Altmann and Trafton have proposed an account of PCE within the AGM model. Although the PCE account is far from satisfactory, as discussed in Chapters 2 and 4, the attempt suggests a direct relevance to the current research agenda. The essential component of the PCE account is the notion of task execution in routine procedural tasks. Grounded in the priming constraint, the AGM account of PCE specifies a mechanism of rote associative learning in procedural tasks. Rote associative learning describes the cueing mechanism between consecutive steps in a learned procedural task: the execution of a task step acts as a cue to prime the execution of the following task step in sequence. In a task involving a PC step, the execution of the preceding step before the PC action comes to serve as a cue to prime the execution of the PC step. Therefore, the AGM account suggests that the occurrence of PCE in a procedural task is minimal under usual circumstances: for instance, no high working memory load.

The notion of associative cueing between procedural task steps has a direct implication for the effect of interruption position. Recent studies examining the effect of interruption position, as described earlier, are concerned with the immediate disruptive effect in terms of resumption time. However, the current investigation is concerned with the effect of interruption position on PCE. The execution of the PC step is, by definition, the final action in a PC task; whether an interruption occurring early in a task has any residual effect on the error remains an open question. In the PM literature, the findings seem to suggest that an interruption might have a residual effect some time later after its occurrence. Therefore, more specifically, the research question is "Is there any position in a PC task where an interruption occurs that is more likely to give rise to PCE?". The AGM account makes predictions regarding the research question; since associative cueing occurs between procedural task steps, as a result, interruptions occurring early in the task should not increase the likelihood of PCE as long as the step preceding the PC action is executed. However, if associative cueing between the PC step and its preceding step is disrupted, the PC step is more likely to be forgotten, resulting in PCE. In other words, interruptions occurring just-before the PC step are more likely to give rise to PCE than interruptions occurring earlier in the task.

Another dimension of interruption that is of interest to the current investigation is the duration of an interruption. The gradual decay process of the AGM model suggests that the disruptiveness of an interruption depends on the duration of the interruption. An interruption has to be cognitively demanding enough to prevent rehearsal of a to-be-resumed goal in a primary task, and the interruption has to last long enough for the goal activation to undergo substantial decay. Therefore, according to the AGM account of PCE, the occurrence of an error not only depends on where an interruption occurs in a PC task but also depends on the duration of the interrupting activity. The series of experiments presented in the next chapter investigate the effect of interruption position and duration on the occurrence of PCE.

Chapter 6

The effect of interruption on post-completion error

Overview

The series of experiments in this chapter investigate the effect of interruption on PCE. Specifically, two dimensions of interruption are being examined: where an interruption occurs in relation to a primary task (interruption position), and how long the interrupting activity lasts (interruption duration). For the purpose of the current set of studies, an interruption is taken to mean the abrupt onset of a different task activity during the execution of a primary task. The experimental methodology used in the current experiments adopts a procedural task paradigm in which the primary task involves participants following and executing a set of pre-defined instructions in a task environment.

A series of four experiments was carried out to examine the effect of interruption position on PCE. The first experiment served as a preliminary study to set up an appropriate task and experimental paradigm to generate the error in a laboratory setting. The primary task is a game-like procedural task in which the participants make doughnuts using a fictional doughnut machine. Consideration of an interrupting task was also taken into account. This first experiment failed to generate enough PCEs to allow for further investigation into the effect of interruption position. The primary task was analysed and compared to Byrne & Bovair's Phaser task, and a number of characteristics were identified as contributing to the failure to provoke the error. The identified characteristics include (1) the presence of a competing signal, (2) the prominence of the PC step button, and (3) the issue of a less practised PC step.

The second experiment involved making modifications to the primary task addressing, specifically, characteristics (1) and (2). The modified primary task was able to generate a high enough level of error and results of the effect of interruption positions were obtained. The interruption task was a complex mental arithmetic task, which lasted 75 seconds. The effect of interruption position suggests that PCEs are more likely to occur after an immediately preceding interruption than an interruption occurring earlier in the task or no interruption at all.

The third experiment investigated the effect of interruption duration on PCE and replicated the effect of interruption position. Two shorter interruption durations, 45 seconds and 15 seconds, were used in the experiment. Results suggest that PCEs are sensitive to the disruptive effect of an interruption as brief as 15 seconds.

The fourth experiment involved modifying the PC action so that it only required execution in some trials. A consistent result of interruption position effect was produced; moreover, an increase in the overall PCE rate was obtained. The increase in PCE rate suggests that the nature of the less frequently practised PC step might have added extra memory demand to the participants.

Experiment 6a: a preliminary study

1 Introduction

The task paradigm adopted in this study is similar to Byrne & Bovair's (1997) — the primary task is a computer-based fictional game, it is a procedural task consisting of a set of pre-defined procedures to be followed in order to complete the task. The current set of experiments could adopt the Phaser task (a game-like task designed by Byrne and Bovair to study PCE) as the primary task to study the effect of interruption on PCE. However, experimental studies on the error phenomenon are scant and it was decided that it would be a contribution to generate the error using a different task environment with similar underlying principles to the Phaser task, namely a procedural task with a PC step to be carried out after task completion. Details of the primary task are described in the next section.

An interrupting task needs to be cognitively demanding to ensure that it would cause disruption to the performance on the primary task. In order to assess this, two types of interruption need to be designed; a complex, cognitively demanding type on the one hand and a simple, less cognitively demanding type on the other, and their disruptive effects to the primary task compared. The cognitively demanding interruption is a complex mental arithmetic task, which lasts for 75 seconds. The simple interruption is a perceptual-motor task, which involves locating and clicking on a target appearing on a computer screen. The next section describes more detail of these interruption tasks.

This first experiment has two objectives: first, using the proposed primary task and interrupting tasks to generate a high enough PCE rate for statistical comparisons. Second, to test the effect of interruption position on PCE occurrences. The AGM model (Altmann & Trafton, 2002) is used as a theoretical framework to make predictions about the effect of interruption position on the error in procedural task execution.

The AGM model suggests that task steps in a learned procedural skill can be viewed as a sequence of associative links: each action step acting as a retrieval cue for the next. This procedural cueing mechanism explains how PCE is usually avoided; people usually manage to carry out PC tasks, such as photocopying, without committing the PCE most of the time. However, if associative cueing between the PC step and its preceding step is disrupted, the PC step is more likely to be forgotten, resulting in PCE. The following hypothesis is proposed: interruptions occurring *just-before* the PC step are more likely to give rise to PCEs than interruptions occurring earlier in a procedural task or no interruption at all.

2 Method

2.1 Tasks

Two computer-based tasks were designed for the current study. The primary task was a procedural task in which participants were required to operate a doughnut-making machine following a set of procedures. The secondary (interrupting) task was a mental arithmetic task.

2.1.1 Primary task — The Wicket Doughnut Machine

The Wicket Doughnut Machine is a procedural task in which participants are required to carry out a set of predefined procedures to operate the machine correctly in order to produce a required number of doughnuts. Figure 6.1 is a screen shot of the Wicket Doughnut machine:

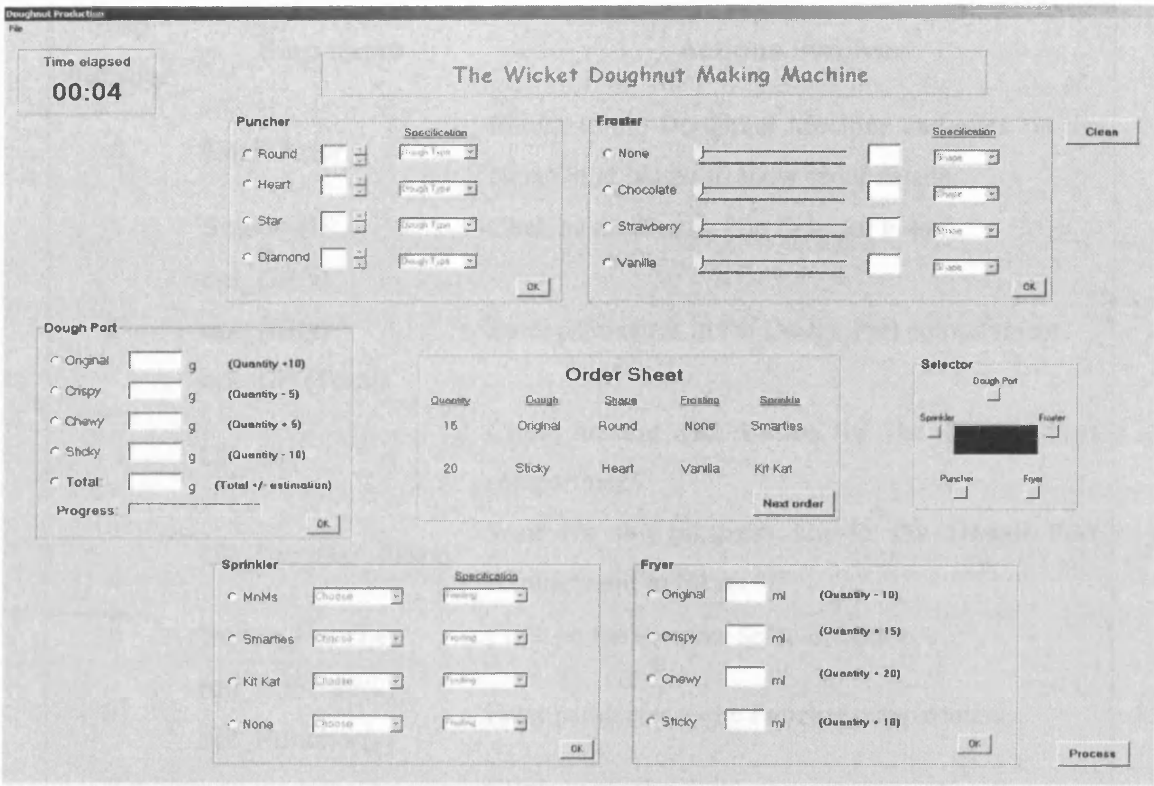


Figure 6.1: The Wicket Doughnut Machine.

The Order Sheet in the centre contains information about how many and what kinds of doughnuts are to be made in each order. Two sets of doughnuts are to be made every time in one order. To produce the doughnuts as specified on the Order Sheet a participants have to operate the five different compartments in the machine in a specific sequence. The compartments need to be operated in the following sequence: 1) Dough Port, 2) Puncher, 3) Froster, 4) Sprinkler and 5) Fryer. Each compartment needs to be activated, using the Selector on the right, before any parameters can be entered. Figure 6.2 summarises all the steps required to operate the machine in a trial.

	Step number	Step name	Actions involved
	0	Next Order	Return to the Doughnut Machine and click on the Next Order button to show order details.
	1	Select_DP	Click on the Dough Port Selector button.
	2[†]	opt_DP(x) opt_DP(y) opt_DP (Total)	Enter parameters in the Dough Port compartment.
	3	DP_OK	Click on the OK button in the Dough Port compartment.
	4	DP_ProgBar_Filling	Wait for the progress bar in the Dough Port compartment to fill up.
Interruption P →	5	Select_Puncher	Click on the Puncher Selector button.
	6[†]	opt_Puncher(x) opt_Puncher(y)	Enter parameter in the Puncher compartment.
	7	Puncher_OK	Click on the OK button in the Puncher compartment.
Interruption Q →	8	Select_Froster	Click on the Froster Selector button.
	9[†]	opt_Froster(x) opt_Froster(y)	Enter parameter in the Froster compartment.
	10	Froster_OK	Click on the OK button in the Froster compartment.
Interruption R →	11	Select_Sprinkler	Click on the Sprinkler Selector button.
	12[†]	Opt_Sprinkler(x) Opt_Sprinkler(y)	Enter parameter in the Sprinkler compartment.
	13	Sprinkler_OK	Click on the OK button in the Sprinkler compartment.
Interruption S →	14	Select_Fryer	Click on the Fryer Selector button.
	15[†]	opt_Fryer(x) opt_Fryer(x)	Enter parameter in the Fryer compartment.
	16	Fryer_OK	Click on the OK button in the Fryer compartment.
Interruption T →			

[†] These steps have different number of parameters to enter: step 2 has three, step 6, 9 12 & 15 have two each. The different number of parameters to enter is a by-product of the cover story of the task.

	17	Process	Click on the Process button.
	18[††]	Report_OK	Wait for the doughnut report to appear, read it and click on the OK button to dismiss it.
Interruption Z →	19	Clean	Click on the Clean button. (This is the PC step)

Figure 6.2: Details of the steps required to operate the Doughnut Machine.

With each of the compartments, participants need to click OK when finish selecting or typing in various parameters. Once the OK button is clicked, information just selected or typed in that compartment are erased and its appearance returns to its beginning state. This ensures that there are no cues in the task environment to distinguish where one is at in the doughnut-making process. After finishing entering various parameters into the five compartments, participants need to click Process (bottom right) to obtain a report indicating how many doughnuts are made. Figure 6.3 is a screen shot showing a completion report.

[††] This step does not afford an error to occur because it involves clicking a modal dialog box.

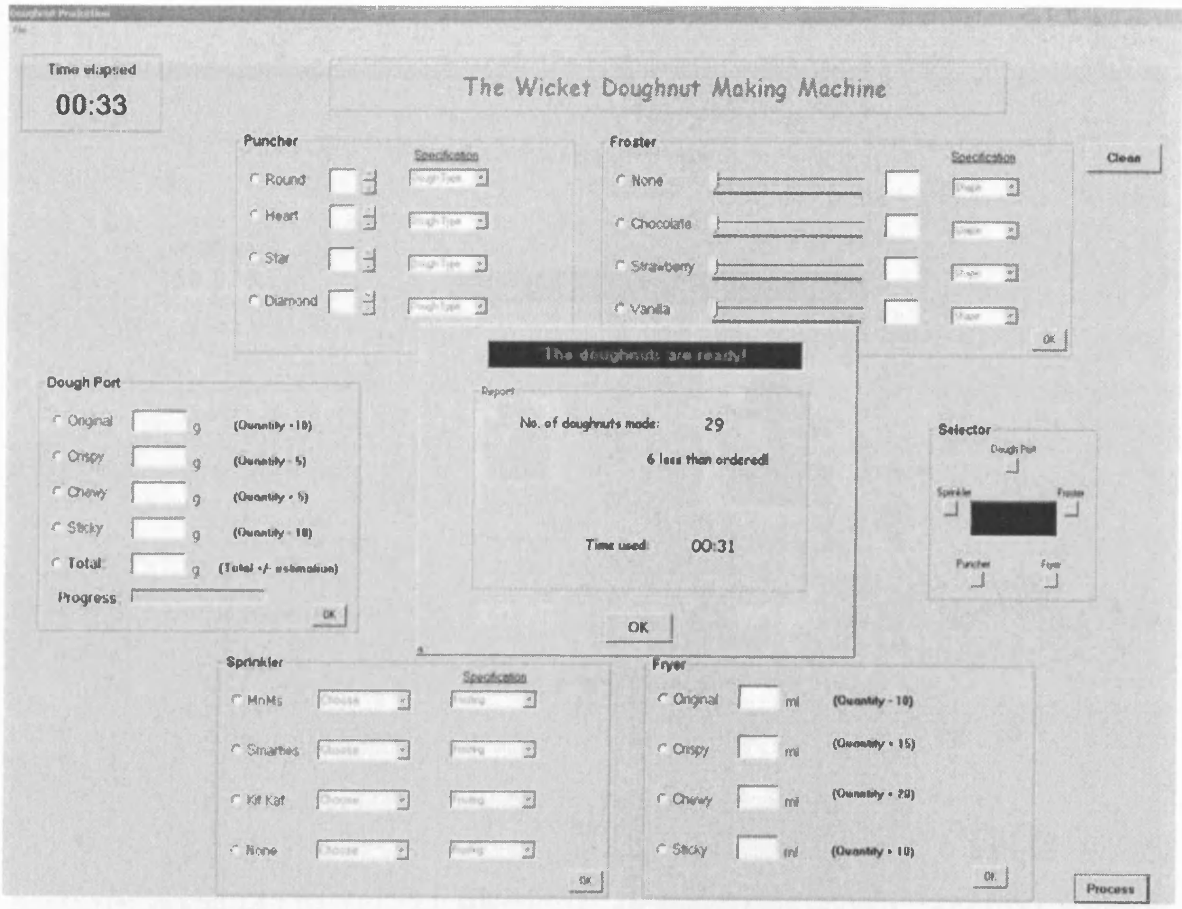


Figure 6.3: A completion report when the task is completed.

When finish reading the completion report participants have to dismiss the report by clicking OK and the screen returns to what it looks like at the beginning of the trial (see Figure 6.1). The final step is then to clean the machine by clicking the Clean button on the top right. Forgetting to execute this step of cleaning the machine is counted as a PCE. To begin the next trial participants have to get the next order by clicking the Next Order button on the Order Sheet.

2.1.2 Complex interrupting task — Doughnut Packaging Express

The complex interruption used in this study is a mental arithmetic task. The task is to pack a certain number of doughnuts following some simple rules. Figure 6.4 shows that there are 52 doughnuts to be packed using two boxes of different sizes, in this case the smaller one can hold 4 doughnuts and the bigger one can hold 9 doughnuts.

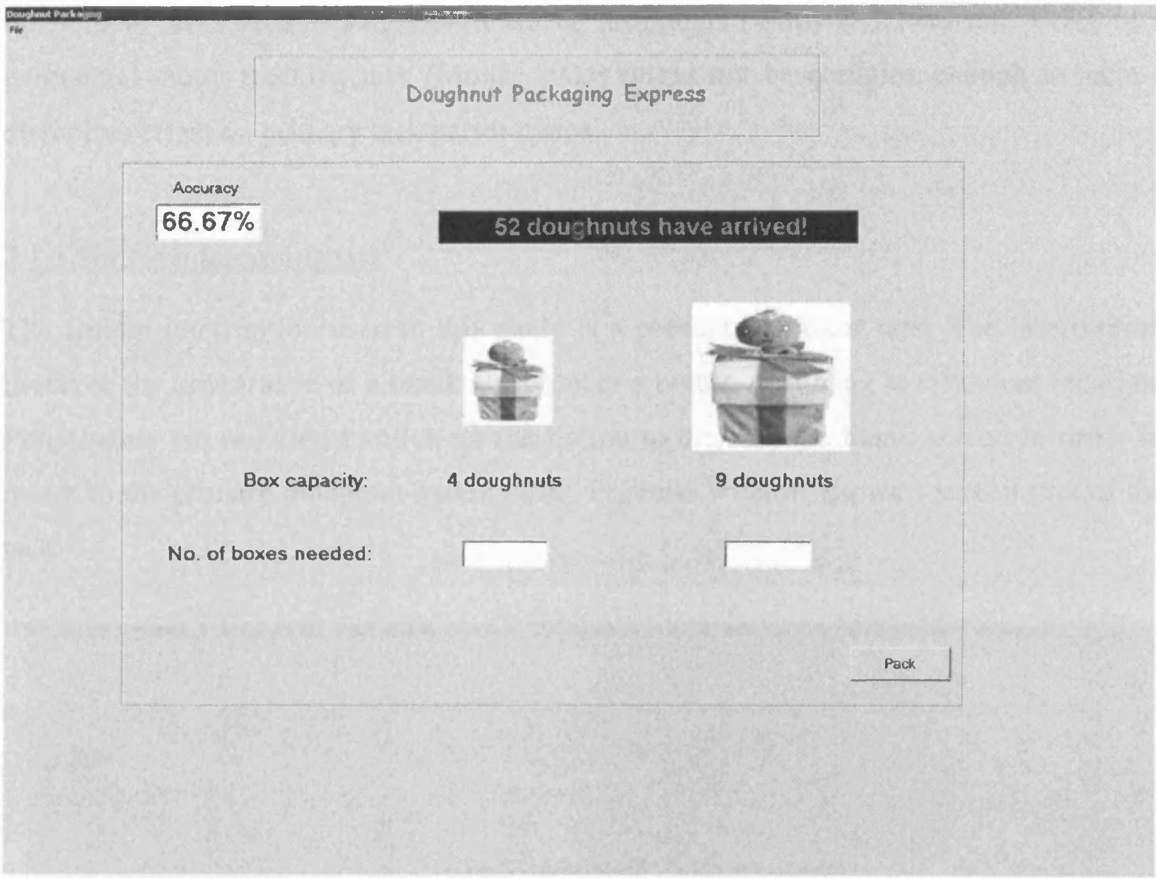


Figure 6.4: The Doughnut Packaging Express.

The task is to calculate how many boxes one might need to pack all 52 doughnuts. In this example, participants might use 13 of the smaller box or 4 of each box type. However, using 6 of the bigger box (which packs 54 doughnuts) results in an incorrect answer since it is required to pack precisely the number of doughnuts as requested. The accuracy score increases when participants get a trial correct but decreases with an incorrect trial.

When an interruption occurs at any point during the doughnut-making task, the computer screen switches to the doughnut-packing task. Participants then have to perform as many packing trials as possible until the screen switches back to the doughnut-making task. After the interruption, participants have to continue the doughnut-making task from where one left off.

The design of this complex mental arithmetic as the interrupting task is based on findings that simple tasks such as single-digit mental arithmetic (Gillie & Broadbent, 1989) or a perceptual-motor tracking task (Monk, 2004) might not be complex enough to have a disruptive effect on primary task performance.

2.1.3 Simple interrupting task

The simple interruption used in this study is a perceptual-motor task. The interruption involves the appearance of a blank screen with a button occurring at a random location. Participants are required to click on the button to dismiss the blank screen in order to return to the primary doughnut-making task. Figure 6.5 below shows a screen shot of the task.

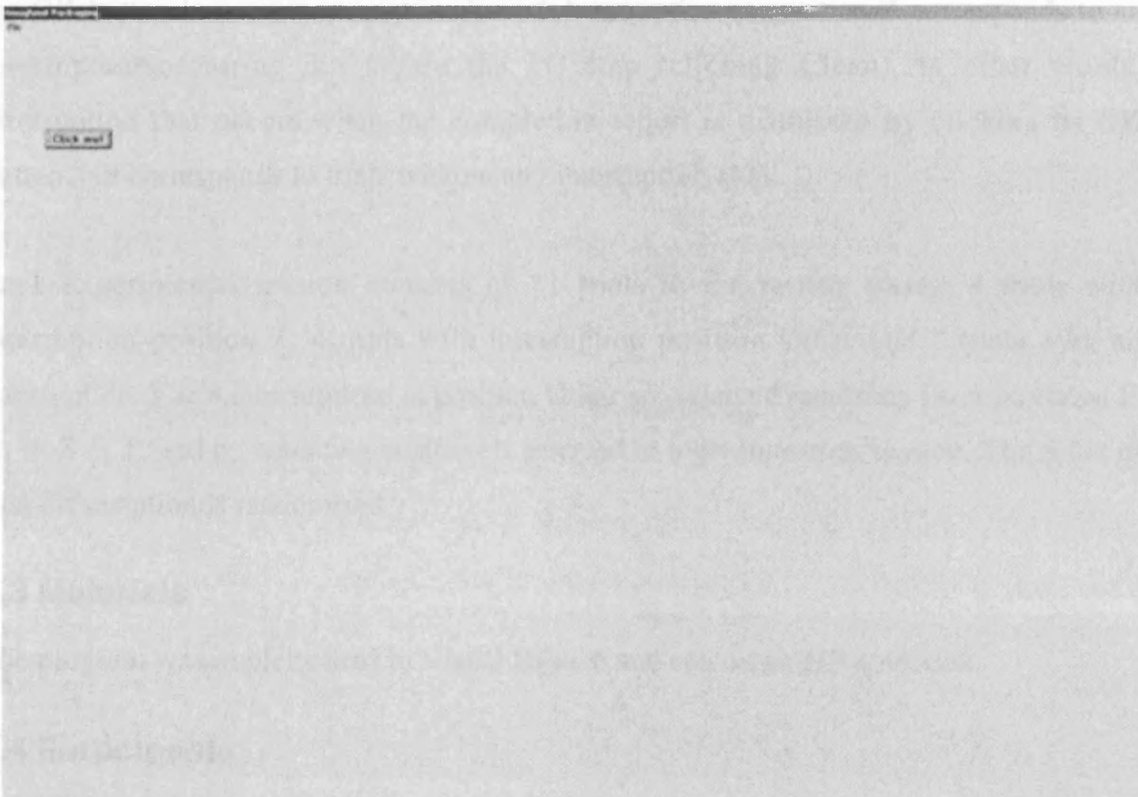


Figure 6.5: The simple perceptual-motor interrupting task.

2.2 Design

The experiment is a 2 × 3 mixed factorial design. The between-subject factor is interruption type which has two levels: simple and complex. Participants in the complex condition have the Doughnut Packaging Express task for 75 seconds as an interruption, whereas participants in the simple condition have the perceptual-motor task as as interruption. The inclusion of the simple interruption condition is expected to have minimal disruptive effect since clicking on an on-screen button requires considerably less cognitive effort than a 75-second mental arithmetic task. The simple interruption has no fixed duration as it disappears whenever the button is clicked.

The within-subject factor (interruption position) has three levels: Just-Before (position Z – see figure 6.2), Other (position P, Q, R, S or T – see figure 6.2) and Nil (no interruptions). Interruption at position P, Q, R, S or T corresponds to an interruption occurring after operating on the Dough Port, the Puncher, the Froster, the Sprinkler or the Fryer compartment respectively. It should be noted that an interruption only occurs when the OK button in a compartment is clicked. Interruption at position Z corresponds to an interruption occurring *just before* the PC step (clicking Clean), in other words, interruption that occurs when the completion report is dismissed by clicking its OK button. Nil corresponds to trials without any interruption at all.

Each experimental session consists of 11 trials in the testing phase; 4 trials with interruption position Z, 4 trials with interruption position Other and 3 trials with no interruption. The 4 interruptions at position Other are selected randomly from positions P, Q, R, S or T, and no repeating position is selected in a given testing session. The order of trial presentation is randomised.

2.3 Materials

The program was implemented in Visual Basic 6 and run on an HP notebook.

2.4 Participants

20 participants, either university undergraduates or postgraduate students, took part in this study. Ages ranged from 21 to 36 with a mean of 27.4. There were 9 female and 11 male. All participants were paid £6 for their participation.

2.5 Procedure

Participants were randomly allocated to one of the two interruption conditions: simple or complex. The participant was then asked to read some introductory documents describing the experiment (Appendix D). This was then followed by a demonstration phase where participants in the 75s-interruption condition observed the experimenter performing both the doughnut-making and the doughnut-packing task separately first. In the training phase, the participant was given two training trials on the doughnut-making task: one with and one without the interrupting task. Any errors occurring in this training phase result in on-screen warning messages and beeps; participants were required to identify and correct the error in order to continue. Participants in the simple condition received the same training procedure except they did not have training on the doughnut-packing task but on the simple perceptual-motor task as the interrupting task. The experimenter was present in the room with the participant during training, and the participant was allowed to ask questions about the tasks if necessary.

Before moving on to the testing phase, participants were encouraged to perform as quickly and as accurately as possible in both the doughnut-making and doughnut-packing task. It should be pointed out that although participants were instructed to produce the correct number of doughnuts by successively guessing the dough "estimation" in operating the Dough Port, the actual number of doughnuts being produced in a trial is random. This is intended to distract the participants away from the real purpose of the experiment.

In the testing phase, any error made results in a beep, but no on-screen message, to remind participants that they have made an error. Each participant was required to perform 11 trials and the experimenter left the participant to carry on the session alone at this point. The entire experiment took approximately fifty minutes.

2.6 Measures

The dependent measure that is of primary interest is the number of PCEs. Two other dependent measures were also taken: a) task resumption error which measures whether the primary task was resumed correctly or not immediately after an interruption, and b) task resumption time which measures the time it takes to resume the primary task after an interruption, i.e. the time difference between the end of an interruption and the first mouse click made. These other measures were taken because, as previous studies suggest (e.g Altmann & Trafton, 2004), they are able to measure any disruptive effect on primary task performance locally. Global measures such as overall task completion time might not be an accurate measure of disruptiveness of the interruption as discussed earlier.

3 Results

3.1 The effect of interruption on PCE rate

Across the two interruption conditions there were only 3 PCEs, 2 of which were committed by the same participant in the complex condition, both after interruption position Z. The remaining PCE was observed in the simple condition and it occurred after interruption position Other. The overall low PCE rate obtained does not support any reliable comparisons between the two interruption conditions and the different interruption positions. However, the data showed that 2 out of 3 PCE occurrences were obtained after the cognitively demanding interruption at position Z.

Since statistical comparisons in interruption conditions and positions were made impractical by the low PCE rate generated, it is necessary to look at other dependent measures in order to establish whether the two interruption conditions have any differential effect in terms of task performance. It would seem pointless to adopt the current experimental paradigm, using the same cognitively demanding interruption, for future investigations if the interruption itself did not have any disruptive effect on the performance of the primary task.

3.2 Differential effects between the two types of interruptions

The first other primary task performance measure to look at was task resumption error after an interruption. In the current context, a task resumption error is defined as an incorrect step executed immediately after an interruption. If there were three consecutive incorrect actions occurring at the same step immediately after an interruption, it only counts as one task resumption error. Therefore, after an interruption each participant would only get a count of 0 or 1 for the error(s). This scheme of error count is to characterise error in terms of the number of incorrect task step executions rather than the number of incorrect actions made in a single task step. This is to eliminate erroneous actions due to trial-and-error behaviour. The same error counting scheme has been adopted by Byrne and Davies (in press).

Since each participant had 8 interruptions in total (4 with position Z and 4 with position Other), a score of a minimum of 0 and a maximum of 8 was obtained for each participant. Furthermore, a task resumption error after interruption Z is operationalised differently to a PCE; a resumption error is any incorrect step made except getting the next order (clicking Next Order), whereas a PCE is a click on Next Order before clicking the Clean button.

In each condition, across participants and trials, there were 80 interruptions. There were 5 task resumption errors (6% error frequency) in the simple condition and 11 task resumption errors (13% error frequency) in the complex condition. A resumption error proportion score was calculated for each participant. The proportion score was obtained by the number of errors divided by the number of interruptions each participant received (8 in this case). An independent-sample t-test showed that there was no reliable differences in the error proportions scores between conditions complex (Mean = 0.14, S.D. = 0.11) and simple (Mean = 0.06, S.D. = 0.12), $t(18) = 1.45$, n.s.. However, the trend of the data indicates that there were more resumption errors in the complex condition than the simple condition.

The results suggest that there were no differences in terms of resumption errors between the two interruption conditions; however, that does not necessarily suggest that the two interruption types do not differ in disruptiveness in terms of other measures. As Trafton et al. (2003) suggest, when other more sensitive measures are looked at the disruptiveness of an interruption is then revealed. The next section looks at the measure of task resumption times.

3.3 The effect of interruption on task resumption times

Another measure of primary task performance is task resumption times. Resumption times were taken for correct performance only, i.e. only resumption times from the *correct* first mouse click after an interruption were included in the analysis. Resumption times for incorrect performance might contain trial-and-error behaviour and, therefore, do not accurately reflect a participant genuinely trying to remember what he/she was about to do. These resumption times for incorrect performance were not included for comparison.

Mean task resumption times were calculated for each participant according to the different task steps: PC step and non-PC step. Because only correct task resumption data were included, the mean resumption times calculated for each participant were not necessarily an average of 4 (4 trials for each of the two task steps) but were fewer in some cases if there were incorrect resumption times.

The mean task resumption times violated the assumptions of homogeneity of variance and covariance; as a consequence, a logarithmic transformation was performed prior to submitting the data to a 2 × 2 (interruption condition × task step) mixed ANOVA. In the analysis that follows, summary statistics, such as the mean and variability, are reported using the geometric means (GM; antilog of the transformed means) and the interquartile range (IR) of the transformed data respectively. Means and standard deviations of the untransformed data are provided in footnotes. All statistical tests were conducted using the transformed data.

Table 6.1 shows the GM and IR of the transformed resumption time data for the two interruption positions.

	PC step	Non-PC step
Simple interruption	1.48s (0.71)	2.32s (1.33)
Complex interruption	4.95s (2.92)	7.30s (5.51)

Table 6.1: GM and IR (in bracket) of the transformed resumption time data according to PC step and non-PC step[3].

A 2 × 2 mixed ANOVA yielded a significant main effect of the within-subject factor; task step ($F(1, 18) = 23.171, p < .001$, Eta squared = .563), and a significant main effect of the between-subject factor; interruption condition ($F(1, 18) = 39.283, p < .001$, Eta squared = .686). The interaction was found to be not significant, $F(1, 18) = .124, p = .729$[4]. Post hoc comparisons using paired-sample t-tests with Bonferroni correction yielded a reliable difference between PC step and non-PC step in both the simple condition and the complex condition.

However, the resumptions at the two different task steps involve clicking on buttons of different locations and sizes on the computer screen. Different positions of the buttons require different physical movement times, which might contribute to the differences in resumption times detected. Therefore, Fitts' Law (Fitts, 1954) was used to approximate the physical movement times for resuming the PC step and any of the non-PC steps (i.e. the "Selector" steps) after an interruption. This is meant to be a rough approximation, so the distance of the movement measured is based on the assumption that the starting position is the location of the "Pack" button of the interrupting Doughnut Packing task. The calculation is not feasible for the simple condition because its interrupting task involves clicking an on-screen button occurring at random locations. This makes

[3] Summary statistics from the untransformed data: in the simple condition, mean task resumption times were shorter for PC step (M = 1.60s, SD = 0.74) than non-PC step (M = 2.50s, SD = 0.99). The same pattern was observed in the complex condition with mean resumption times PC step (M = 5.56s, SD = 3.16) shorter than non-PC step (M = 8.33s, SD = 4.60).

[4] The same analysis on the untransformed data yielded a very similar pattern of results; both main effects were found significant; the within-subject task step ($F(1, 18) = 19.67, p < 0.001$) and the between-subject interruption condition ($F(1, 18) = 16.4, p = 0.001$). The interaction was found significant ($F(1, 18) = 5.07, p = 0.037$).

assuming a reasonable starting position of the movement impractical. So physical movement times are calculated for the complex condition only.

Fitts' Law yielded movement times of 0.5 sec and 0.67 sec for PC step and non-PC step respectively (see Appendix E for more details of the calculations). These movement times were taken out from the untransformed resumption times in the complex condition accordingly and the data was subjected to the same transformation and analysis as before. Table 6.2 below shows the GM and IR of the transformed resumption time data, in the complex condition, after taken out the corresponding movement times.

	PC step	Non-PC step
Complex	4.41s (2.92)	6.53s (5.51)

Table 6.2: GM and IR (in bracket) of the transformed resumption time data with movement times taken out according to different interruption positions[5].

Resumption times, with movement times removed, in the complex condition were analysed together with resumption times in the simple condition. A 2 × 2 mixed ANOVA yielded the same pattern of results as earlier. Both main effects were significant with large effect sizes: task step ($F(1, 18) = 21.616$, $p < .001$, Eta squared = .546) and interruption condition ($F(1, 18) = 27.471$, $p < .001$, Eta squared = .604). The interaction was found to be not significant, $F(1, 18) = .091$, $p = .766$. Post hoc comparisons, with Bonferroni correction, comparing PC step and non-PC step in the complex condition yielded a reliable difference. Planned contrasts also revealed reliable differences between complex and simple condition with respect to PC step ($t(18) = 5.235$, $p < .001$) and non-PC step ($t(18) = 4.405$, $p < .001$).

5 Summary statistics from the untransformed data, with movement times taken out, in the complex condition: PC step (M = 5.06s, SD = 3.16) and non-PC step (M = 7.66s, SD = 4.60).

4 Discussion

The original hypothesis predicting a significant increase in PCEs with a cognitively demanding interruption (a mental arithmetic task) *just before* the PC step (position Z) is not supported by the observed results. Overall there were only 3 PCE occurrences, with 2 committed by the same participant in the experimental condition and 1 occurring in the control condition. Interestingly, consistent with the prediction, the 2 PCEs in the experimental condition occurred after interruption position Z and not other positions. However, the overall low PCE rate makes investigation of the interruption position effect impractical.

The cognitively demanding interruption was found to have a disruptive effect on primary task performance. Although no significant difference was detected in terms of task resumption errors between the complex and simple conditions, the trend of the data suggests that there were more resumption errors in the complex condition than the simple.

Results from task resumption times suggest that participants took longer to resume the primary task after a cognitively demanding interruption than a simple perceptual-motor interruption for both PC and non-PC steps. This suggests that the 75s mental arithmetic task was successful in engaging the participants, preventing them from rehearsing where they were in the primary doughnut-making task, and exhibiting a detrimental effect on primary task performance. Although the reliable difference between interruption conditions was based on resumption times in the complex condition having the physical movement time estimated and removed, the same could not be carried out in the simple condition due to the design of the interrupting task. The comparison is still valid and, in fact, it is a more conservative one, because if estimation of the physical movement times were obtainable in the simple condition and removed, the difference between the two conditions would be even larger, giving the same conclusion.

Differences in resumption times were also detected between PC step and non-PC step. Results suggest that it was faster to resume the PC step than to resume one of the five non-PC task steps. The same result was obtained even when the approximate physical movement times were removed in the complex condition.

Similar results were found on a study investigating the effect of interruption positions on programming a VCR (Monk, Boehm-Davis & Trafton, 2004). It was found that resumption times were longer for mid-subtask interruptions than interruptions before a new subtask or during a repetitive procedure like scrolling. The difference is explained in terms of the AGM model that different task or subtask points impose a greater encoding demand. The longer resumption times is due to the need to rehearse more information about the goal state and its links to the next step. Extending the explanation to our current finding, it is possible that because after interruption Other (which happened at position P, Q, R, S or T) the task still remains incomplete, in order to resume the next correct step, the cognitive system needs to successfully encode/differentiate between the different states of the five similar subtasks. On the other hand, after interruption Z the task has already been completed (doughnuts were made) so the encoding of the goal state might be relatively less demanding. This is because the PC step is "standalone" in the sense that it has no similar subtasks to differentiate with, resulting in shorter resumption times. Although the same difference in interruption position was also present in the simple condition, the result is less conclusive. This is because the corresponding physical movement times could not be obtained and, thus, not removed. As a result, the observed difference in this condition might be confounded.

Using task resumption time as a behavioural measure (Trafton et al, 2003), a differential disruptive effect between the two interruption conditions was observed in the current experiment. The difference observed in resumption times is also consistent with previous findings showing that even an interruption of 5 seconds results in longer resumption times than a 1 or 1/4 second interruption (Monk & Trafton, 2004). However, the current designs of the two interruption conditions were not to investigate the effect of different interruption duration, but to establish whether the 75-second mental arithmetic task is disruptive to primary task performance as an interruption.

The main objective of the current experiment investigating the effect of interruption position on PCE is not fulfilled because it failed to generate a high enough number of PCEs. The overall low PCE rate, which is 1.4% (3 out of 220 opportunities), generated from the current experiment has not reached the 5% threshold of systematicity.

The current study adopts a similar experimental paradigm to Byrne & Bovair's study utilising a procedural task, which has a predefined sequence of operations that participants have to follow. Participants were trained to perform the task correctly to ensure they have the correct knowledge of the task procedures; this ensures that any errors that arise from the task are procedure-based and not knowledge-based. However, the task used in this experiment differs from Byrne & Bovair's task on a number of dimensions. One might question the validity of the task for investigating PCEs: in other words, whether the null effect of interruption on PCE rate might be attributed to the current task being an unsuitable candidate to generate PCEs. Therefore, the following is a systematic comparison between the current task and the one used in Byrne & Bovair's study. The purpose is to identify if there are any characteristics that the current task might lack which contribute to generating PCEs.

4.1 An analytical comparison and contrast between the doughnut-making task and Byrne & Bovair's Phaser task

A very simple comparison between Byrne & Bovair's task, and the current task, identifies a structural similarity arguing for validity of the current task in studying PCEs. The essence of the comparison is that in both tasks, the execution of the final step occurs after one has acknowledged a completion signal (see Table 6.3). The omission of the final step is then a PCE. Nevertheless differences between the Phaser task and Doughnut task should be considered in more detail and this contrast might shed light on the different level of PCE rate obtained.

	Phaser Task (Byrne & Bovair's task)	Doughnut Task (current task)
Pre-completion steps	A set of fixed procedures	A set of fixed procedures
Completion signal	Decide if the goal has been completed or not (whether the enemy ship had been destroyed or not)	A report of the completed goal (a report of how many doughnuts are made)
Final step of the task	Clicking on "Tracking" button	Clicking on "Clean" button

Table 6.3: A simple descriptive comparison of Byrne & Bovair's task and the current task in terms of task structure.

A closer look at the Phaser and Doughnut-making tasks reveal some important differences as follow:

Presence of a competing signal:

When the completion signal (a doughnut report) occurred in the Doughnut task participants were required to acknowledge the report then dismiss it by clicking an "OK" button on the report. The resulting screen is the same as before the task was completed; in other words, there is no information on screen to differentiate the pre-completion and postcompletion state of the task. In contrast, the completion signal remains on screen once the Phaser task is completed (see arrow 1 in Figure 6.5). Having the completion signal remaining on screen helps one to differentiate the state of the task, and this might "urge" participants to move on to the next step and omit the PC step. This luring effect might be stronger especially when the completion signal includes information such as "Return to main control". The term "competing" signal is used to describe the signal competing against the PC step so that upon completion of a task, one has the choice of executing the PC step or of responding to the competing signal, omitting the PC step. The inclusion of a competing signal upon task completion is an attempt to exploit the notion that when resuming a task after an interruption one uses cues in the external environment to help "reconstruct" the next step in the task sequence.

Prominence of the PC step button:

In the doughnut-making task the PC step is executed by pressing a separate button ("Clean") on the task interface; on the other hand, the PC step (turning off the "Tracker") in the Phaser task is carried out by pressing the "Tracking" button which is also used for turning on the "Tracker" (see arrow 2. in Figure 6.6). So the same button is used for two different functions ("on" and "off") in the Phaser task. The first click to turn the "Tracker" on is analogous to pressing the "Process" button in the Doughnut task, and the second click to switch the "Tracker" off is like pressing the "Clean" button.

Arrow 1. is the completion signal indicating "Romulan vessel destroyed. Return to main control."
Arrow 2. highlights the "Tracking" button.

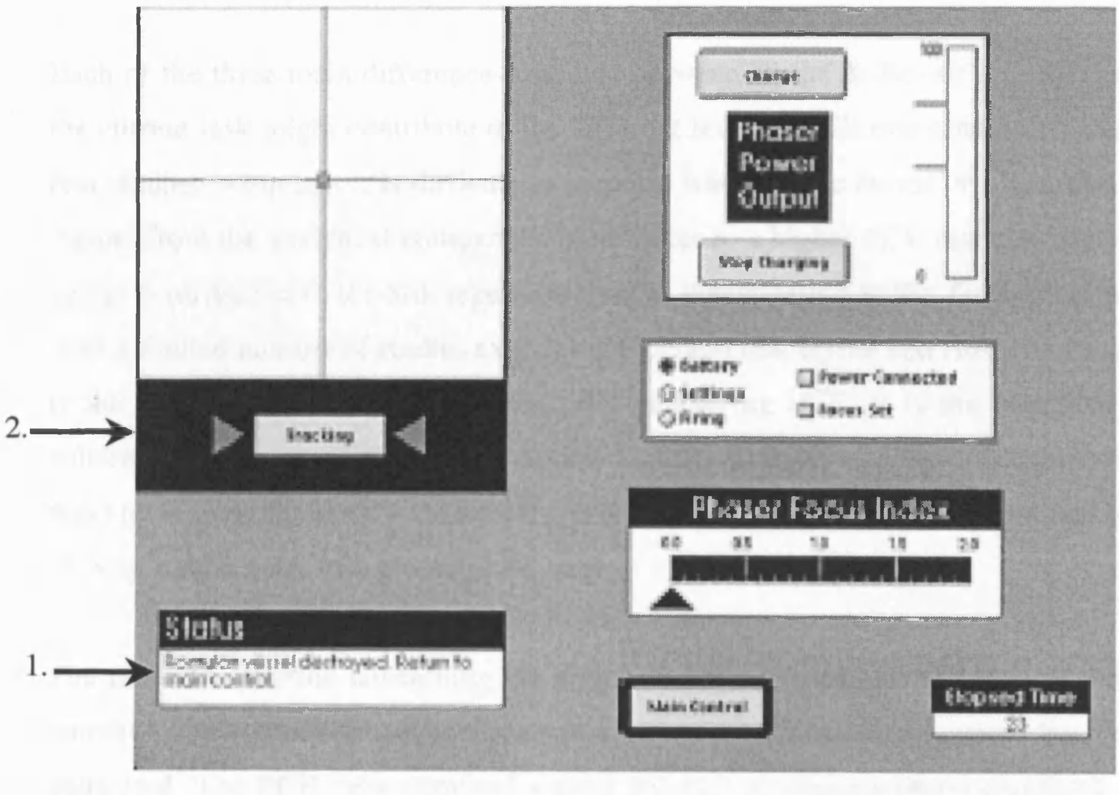


Figure 6.6: A screen shot of the Tactical task (adapted, with permission, from Chung & Byrne (2004) which is the same as the Phaser task in Byrne & Bovair, 1997).

A less practised PC step due to a cyclic task structure:

The completion signal in the Phaser task is essentially a decision point; when the goal has been completed move on to the PC step (clicking "Tracking") of the task, otherwise carry out the previous sequence of steps again. This cyclic (or looping) of the pre-completion steps means that, within a given trial, these steps are executed more often than the PC step. However, this particular property is not present in the Doughnut task as the completion signal is not a decision point but simply a task completion report. Therefore, the pre-completion steps and the PC step in the Doughnut task are practiced equally often within a single trial. The cyclic pre-completion property in the Phaser task might increase the chances of one making the PCE due to the PC step being a less practised step compared to all other steps within the procedure. One could imagine that a participant carrying out this task going through a few cycles of "*Am I done yet? No, do it again.*" When the task is finally done this might give the participant the feeling of "*Ah, now I am done. Let's move on*", and subsequently forgetting about the PC step.

Each of the three main differences outlined between Byrne & Bovair's Phaser task and the current task might contribute to the different level of PCE rate obtained between the two studies. Although it is difficult to pinpoint whether one factor or a combination of factors from the analytical comparison contributes to a higher PCE rate, the analysis is a useful contribution to the task repertoire used in investigating PCEs. Given that there are only a limited number of studies examining PCE and that Byrne and Bovair's Phaser task is the main task paradigm in successfully generating PCE, it is the objective of the subsequent experiments to use the current procedural task paradigm (doughnut-making task) to address the identified factors: presence of a competing signal, prominence of the PC step button and a less practised PC step.

The research question concerning the effect of interruptions on PCE rate set out in the current experiment could not be answered adequately because of the overall low PCE rate generated. The PCE rates obtained were 1.8% (2/110 opportunities) and 0.9% (1/110 opportunities) in the complex interruption condition and the simple interruption condition respectively. The obtained error rates are considerably lower than the 9.3% (13/140 opportunities) generated in Byrne & Bovair's study using the Phaser task with no working memory manipulation. The goal of the next experiment is to redesign the doughnut-making task to attempt to generate a higher PCE rate than the current experiment, so that the effect of interruption position on PCE rate can be assessed.

Experiment 6b: The effect of interruption position

1 Introduction

In retrospect, apart from the three factors identified in the analytical comparison between the doughnut-making task and the Phaser task, there might be a further reason for the low PCE rate obtained in the previous experiment. When the goal of making doughnuts is completed one has to click on a button labelled "Clean" to clean the machine before clicking "Next Order" to begin the next trial. However, to the participants clicking "Next Order" might not be viewed as the first step to start the next trial but clicking the "Clean" button first instead. In order to make it clear to the participants that the "Next Order" button is the beginning of a new trial, a clear separation between trials is needed. In most real-world examples where PCEs could occur, there is a subsequent task awaiting when the main goal is accomplished. For example, forgetting to collect the original after photocopying, one might have a task to distribute the photocopies when they are done; when using an ATM and forgetting to take the cash card back, one might need to go to a restaurant after the withdrawal is done. On reflection, the absence of a task to separate one trial from the next might not resemble situations where PCEs arise. Therefore, one of the changes made to the doughnut-making task in Experiment 6a is to include a relatively small task to separate one trial from the next. Furthermore, this small task should offer a competing signal so that it "urges" one to move on to it upon completion of the main task of making doughnuts.

In the current experiment, the following changes are made to the task paradigm: firstly, a small task is included to separate one trial from the next to address the issue of having a subsequent task to carry out after the main goal is completed. Secondly, in order to address the absence of a competing signal, a competing signal is included upon completion of the doughnut-making task to "urge" the participant to move on to the subsequent task. Thirdly, to address the issue of prominence of the PC step button, the "Clean" button is integrated with the existing "Process" button so that the PC action is not on a separate button.

The effect of interruption position on PCE could then be studied if the changes made to the task paradigm managed to generate an overall error rate for statistical comparisons. The issue of a less practised PC step is addressed in the final study — Experiment 6d.

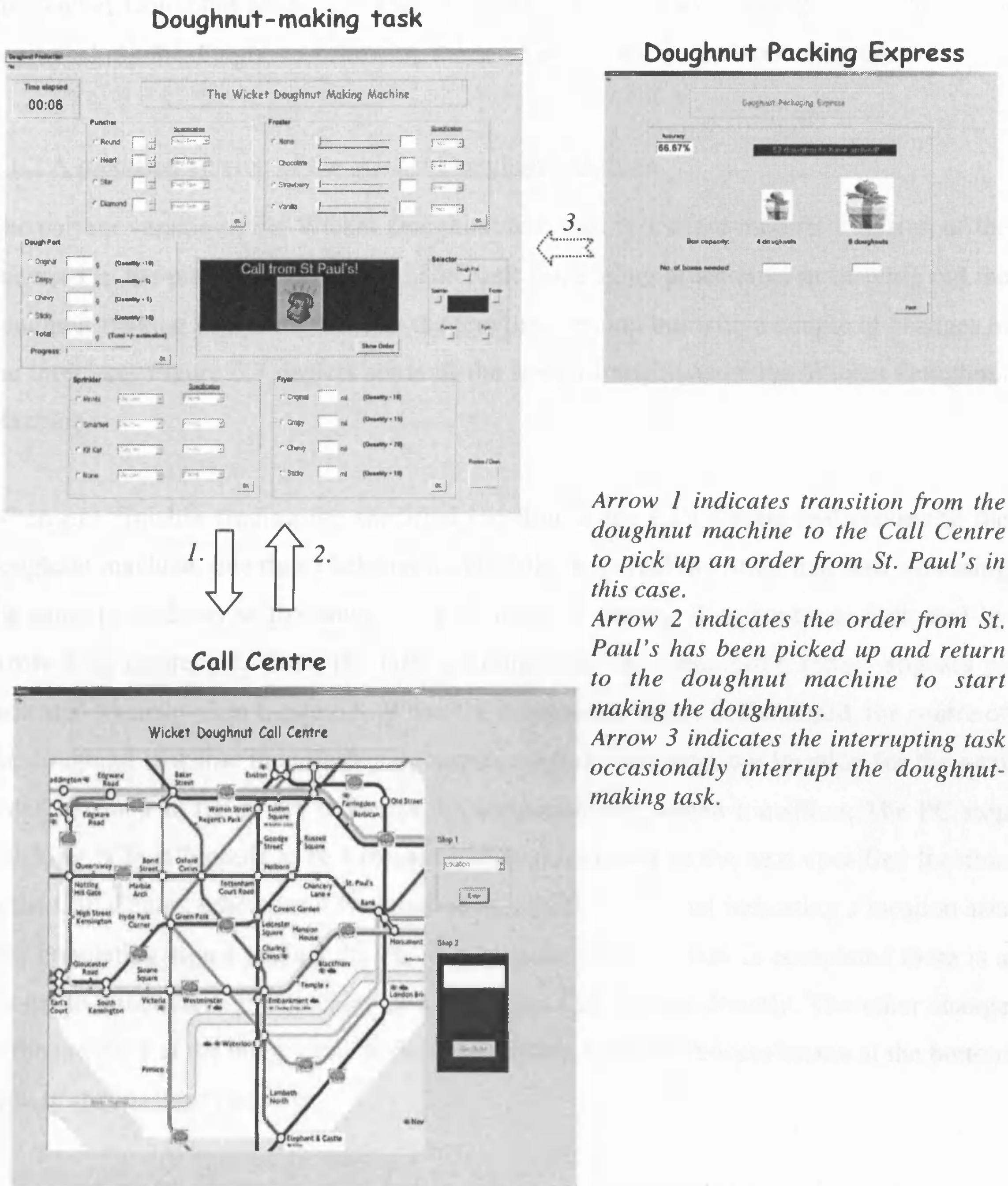
2 Method

2.1 Tasks

In this experiment, an additional task was used apart from the doughnut-making and doughnut-packing tasks. The additional task is a relatively small task and is implemented on a separate computer terminal. Two computer terminals were used in this experiment: the rationale is to simulate an environment where one has to physically move away from the main task (making doughnuts) when it is completed in order to move on to a different task. It was felt that a lot of the real-world PCE situations involve one physically moving away from the artefact once the task is accomplished.

2.1.1 The primary task — Call Centre and Doughnut-making

The Call Centre is a simple search task where one has to find a specified location from the London Underground Map to get an order to make doughnuts. This search task is to be carried out at the beginning of each trial. Figure 6.7 depicts the transition between the modified version of the Doughnut-making task and the Call Centre.



Arrow 1 indicates transition from the doughnut machine to the Call Centre to pick up an order from St. Paul's in this case.
Arrow 2 indicates the order from St. Paul's has been picked up and return to the doughnut machine to start making the doughnuts.
Arrow 3 indicates the interrupting task occasionally interrupt the doughnut-making task.

Figure 6.7: The Wicket Doughnut Machine and the Call Centre.

At the beginning of a trial, the centre of the Wicket Doughnut Machine indicates a location to collect an order (St. Paul's in this case) and the participant is required to physically turn to the Call Centre computer terminal to find the location. Once the location is found and entered following two simple steps, the participant then returned to the Wicket Doughnut Machine terminal to show the order by clicking Show Order and begin making the doughnuts following the same procedure as in Experiment 6a.

2.1.2 A modified version of the Wicket Doughnut Machine

The current version of the Wicket Doughnut Machine is a slight modified version of the one used in the previous experiment. The basic underlying procedures in carrying out the doughnut-making task is the same as the previous version but with a couple of changes to the interface. Figure 6.7 depicts some of the screen transitions of the Wicket Doughnut Machine.

When one finishes finding the specified location in the Call Centre and returns to the doughnut machine, one then clicks on Show Order to reveal the order and start executing the same procedures as the previous experiment in making doughnuts, as indicated by arrow 1 in figure 6.8. Once the task is completed, the completion report appears as indicated by arrow 2 in figure 6.8. When the completion report is dismissed, the centre of the doughnut machine then flashes up another signal indicating the location for the next order, arrow 3 in figure 6.8 indicates the corresponding screen transition. The PC step (clicking "Clean") needs to be executed before responding to the next specified location in the Call Centre, otherwise it is classified as a PCE. The signal indicating a location acts as a competing signal against the PC step because once the task is completed there is a choice to either clean the machine or to go to the Call Centre directly. The other change to the interface is the integration of the Clean button with the Process button at the bottom right of the machine interface.

Figure 6.9 below summarises the high-level steps of the modified task. The task steps involved are very similar to the previous experiments with the addition of the Call Centre at the beginning.

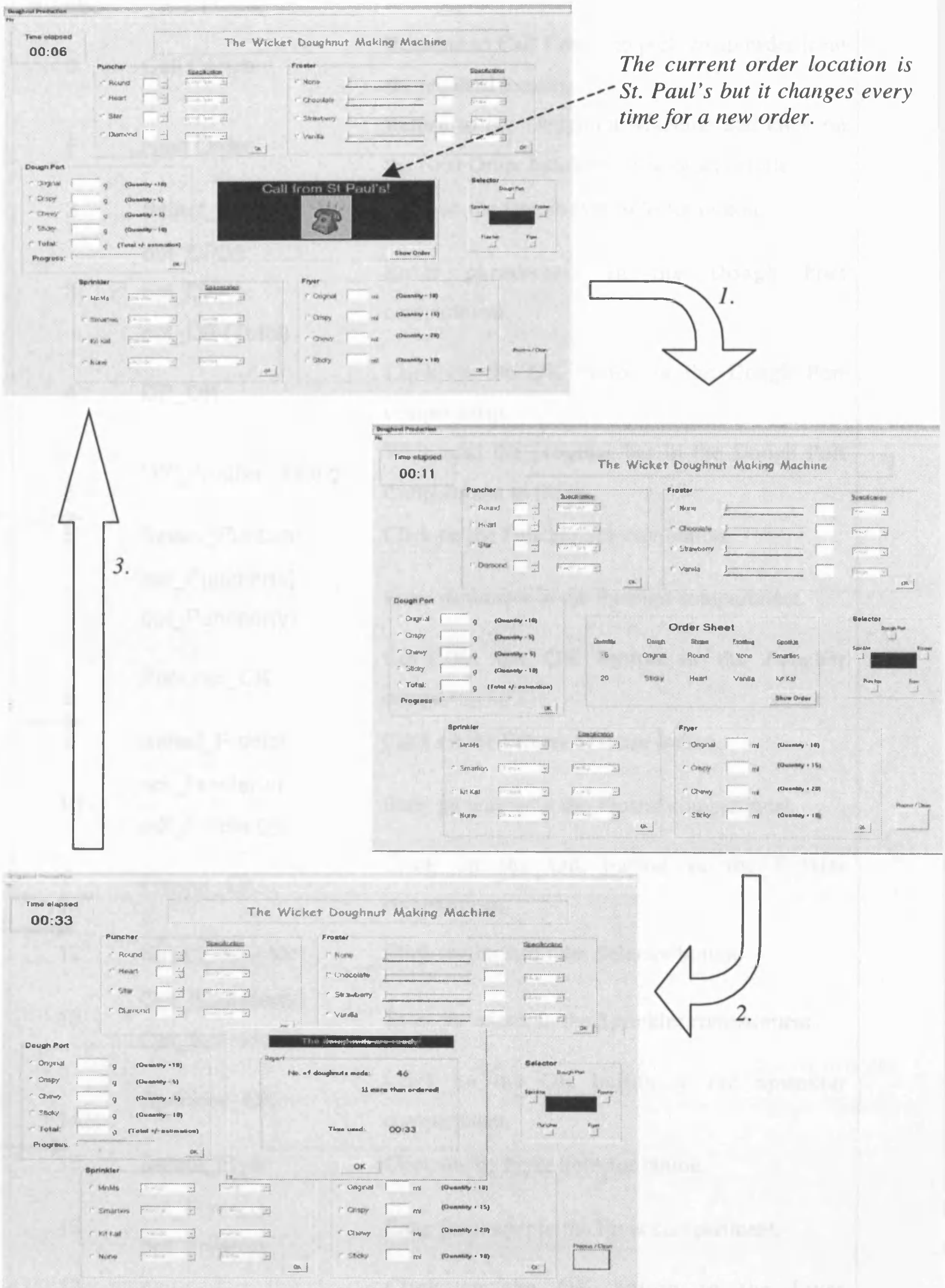


Figure 6.8: The Wicket Doughnut Machine.

	Step number	Step name	Actions involved
	0	Call Centre	Respond to Call Centre to pick up an order from the required location.
	1	Next Order	Return to the Doughnut Machine and click on the Next Order button to show order details.
	2	Select_DP	Click on the Dough Port Selector button.
	3	opt_DP(x) opt_DP(y) opt_DP (Total)	Enter parameters in the Dough Port compartment.
	4	DP_OK	Click on the OK button in the Dough Port compartment.
	5	DP_ProgBar_Filling	Wait until the progress bar in the Dough Port compartment to fill up.
Interruption P →			
	6	Select_Puncher	Click on the Puncher Selector button.
	7	opt_Puncher(x) opt_Puncher(y)	Enter parameter in the Puncher compartment.
	8	Puncher_OK	Click on the OK button in the Puncher compartment.
Interruption Q →			
	9	Select_Froster	Click on the Froster Selector button.
	10	opt_Froster(x) opt_Froster(y)	Enter parameter in the Froster compartment.
	11	Froster_OK	Click on the OK button in the Froster compartment.
Interruption R →			
	12	Select_Sprinkler	Click on the Sprinkler Selector button.
	13	Opt_Sprinkler(x) Opt_Sprinkler(y)	Enter parameter in the Sprinkler compartment.
	14	Sprinkler_OK	Click on the OK button in the Sprinkler compartment.
Interruption S →			
	15	Select_Fryer	Click on the Fryer Selector button.
	16	opt_Fryer(x) opt_Fryer(y)	Enter parameter in the Fryer compartment.
	17	Fryer_OK	Click on the OK button in the Fryer compartment.
Interruption T →			

	18	Process	Click on the Process button.
	19[††]	Report_OK	Wait for the doughnut report to appear, read it and click on the OK button to dismiss it.
Interruption Z →	20	Clean	Click on the Clean button. (PC step)

Figure 6.9: Details of the steps required to operate the Doughnut Machine.

2.1.3 Interrupting task

The interrupting task of packing doughnuts with the Doughnut Packaging Express is identical to the one in the previous experiment.

2.2 Design

The design of the current experiment is identical to the previous experiment except there is no between-subject factor. This is a within-subject design with one independent variable — interruption position — which has three levels: Just-Before (position Z), Other (position P, Q, R, S or T) and None (no interruption). The interruption used is the 75-sec mental arithmetic task.

Each experimental session consists of 11 trials in the testing phase; 4 trials with interruption position Z, 4 trials with interruption position Other and 3 trials with no interruption. The 4 interruptions at position Other are selected randomly from position P, Q, R, S or T, and no repeating position is selected in a given testing session. The order of trial presentation is randomised.

2.3 Materials

The programs were written in Visual Basic 6 and run on two different computer terminals: a HP Desktop and a Compaq desktop. The doughnut machine program and the doughnut-packing program were run on the HP desktop and the Call Centre was run on the Compaq desktop. The two computer terminals were arranged at 90° so that the

[††] This step does not afford an error to occur because it involves clicking a modal dialog box. Therefore, it is not included in calculating the overall error opportunities in the analysis

participant had to turn away from either computer depending which task s/he was carrying out.

2.4 Participants

35 participants, either university undergraduate or postgraduate students, took part in this study. Ages ranged from 19 to 37 with a mean of 24.8. There were 20 female and 15 male. None of the participants had taken part in the previous experiment. All participants were paid £6 for their participations.

2.5 Procedure

The procedures were very similar to the previous experiment. Participants read some documents describing the experiment then went through a demonstration and training phase. Participants first observed the experimenter performing both the doughnut-making and the doughnut-packing task separately first. When performing the doughnut-making task the experimenter explained one needs to respond to the Call Centre when the location signal flashes in the doughnut machine. The experimenter also demonstrated how to respond to a call using the Call Centre and that the doughnut machine needs to be cleaned by clicking the "Clean" button, after each completed order, before responding to the Call Centre. Participants were then given two training trials on the doughnut-making task; one with and one without the interrupting doughnut-packing task. Any errors occurring in this training phase result in on-screen warning messages and beeps; participants were required to identify and correct the error in order to continue. The experimenter was present in the room with the participant during training, and the participant was allowed to ask questions about the tasks if necessary.

In the testing phase, participants were required to perform 11 trials in total and the experimenter left the participant to carry on the session alone at this point. The entire experiment took approximately an hour.

2.6 Measures

Two dependent measures are of primary interest: first, the number of PCEs made; secondly, the time one takes to resume an action after an interruption. Other errors are also of interest and recorded.

3 Results

Data from four participants were excluded in the analysis; two of them were making the PCE on every trial. The pattern of results from these two participants suggests the PCEs made may be due to incorrect knowledge of the task, so that the observed PCEs may not be genuinely procedural errors. One other participant did not follow the task instructions properly and one had a lost data file.

3.1 Overall errors

There were a total of 330 errors across the 31 participants. An error is defined as any incorrect action deviating from the correct sequence. The error count for each participant was carried out in relation to the task steps; for example, if there was more than one error at task step X, the error count is still one for that task step. The same scheme of error counting was used as the previous experiment: characterise error in terms of the number of incorrect task step executions rather than the number of incorrect actions made in a single task step. This is to eliminate erroneous actions due to trial-and-error behaviour.

Over half of the participants (20 out of 31) made at least one PCE (error at "Clean" step) and a total of 56 PCEs were obtained. If the occurrences of errors, in general, were random then each task step should account for 5% (1/20) of the total errors. The obtained number of PCEs account for about 17% of the total errors. A Binomial test suggests the occurrences of PCE is above chance level, $z = 9.98$, $p < .001$.

An alternative way of interpreting the data is assessing the error against its systematicity level (e.g. Byrne & Bovair, 1997, and Payne & Squibb, 1990). Systematicity is defined as a proportion ratio of number of occurrences to the number of opportunities for that error. Since there were 11 trials for each of the 31 participants, this gives a total of 341 opportunities for an error at each task step. Figure 6.10 shows the systematicity, also referred to as error rate, for all task steps[6].

[6] Task step [19] has been omitted from the figure and also from subsequent analysis because it involves only a mouse click on a modal dialog box, which does not afford an opportunity for error occurrence.

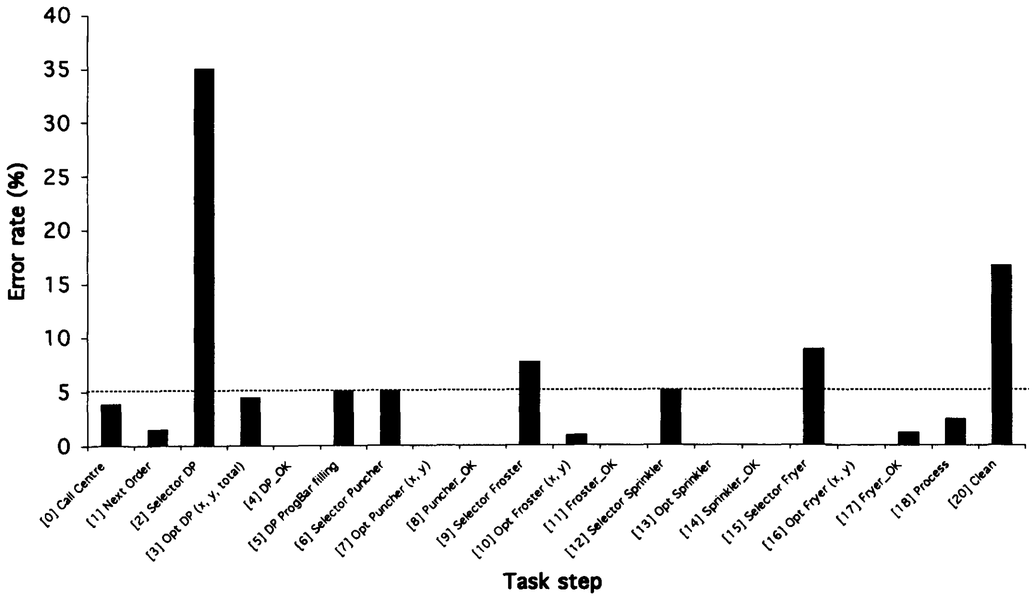


Figure 6.10: Error rate (%) of each task step across trials and participants.

A 5% level of systematicity has been used as an indicator of an error occurring in a systematic fashion (see Byrne & Bovair, 1997). PCE appears to be highly systematic with an error rate of 16%. A number of non-PCEs also appear to be systematic and they mostly occurred at one of the 5 "Selector" steps with "Selector DP" having the highest error rate. The occurrence of errors at these "Selector" steps is discussed later in the chapter. Task step [5] involves waiting for a progress bar to fill up; an error occurs when an action is made before the progress bar is filled up. The occurrences of this error suggest that the participants were being impatient: not waiting long enough for the progress bar to fill up before carrying out the next action. This error is not pursued further in the analysis because its nature is not of particular interest to the current thesis.

3.2 Interruption position effect on PCE

Table 6.4 shows the distribution of the PCE occurrences according to the different interruption positions. Each participant received 4 interruptions at position Z, 4 interruptions at position Other and 3 trials with no interruptions; the total number of opportunities for each interruption position was calculated across all 31 participants.

Error rates (%) of PCE occurrences were calculated for the three different interruption trials for each participant. The error rates are the number of PCEs divided by the number of opportunities for the error. The use of error rates for comparison eliminates biases imposed by the different number of opportunities among the interruption positions.

	Interruption position		
	Z	Other	Nil
Total no. of PCE	37	11	8
(Total no. of opportunities)	(124)	(124)	(93)
Mean error rate	29.8%	8.9%	8.6%
(SD)	(32.5)	(19.9)	(21.0)

SD = Standard deviation

Table 6.4: Number of PCEs and their mean error rates with respect to the different interruption positions.

A one-way repeated ANOVA on the error rates showed a large significant main effect of interruption position. The scores did not conform to the assumption of sphericity, the Greenhouse-Geisser correction was used, $F(1.338, 40.149) = 9.921$ $p = .001$; Eta squared = 0.249. Planned contrasts comparing position Z to position Other showed a reliable difference, $t(30) = 3.297$, $p = .003$, and position Z versus Nil also showed a significant difference, $t(30) = 3.339$, $p = .002$. There was no reliable difference between position Other and Nil, $t(30) = .09$, $p = .929$.

3.3 Interruption position effect on non-PCE

The next question is concerned with whether the observed interruption position effect is unique to PCEs? In other words, does interruption position have the same effect on errors occurring at other non-PC steps in the task? For the purpose of the comparisons, all non-PCEs were categorised into "immediately after interruption", "later after interruption" and "no interruption" (including errors before an interruption). These are equivalent to the PCEs' interruption position Z, Other and Nil respectively.

The calculation of error opportunities and error count for each of the categories is as follows:

3.3.1 "Immediately after interruption"

Error count

An action following immediately after an interruption is counted as an error if it was executed on the incorrect task step.

Error opportunity

Each participant has 4 opportunities for the error from the 5 interruption positions P, Q, R, S or T.

3.3.2 "Later after interruption"

Error count

Any action on the second task step onwards following an interruption is counted as an error if it was carried out on the incorrect task step.

Error opportunity

For example, after an interruption on position P (before Selector Puncher) there are 12 opportunities for a non-PCE. The step before the PC step was excluded as an opportunity because there is no alternative action a participant could perform, which would count as an error, apart from clicking on a modal dialog box.

There were no opportunities for this category of error after interruption T and Z, therefore, their subsequent task steps were not counted as error opportunities. Depending on which 4 of the 5 interruption positions a participant received, each participant might have different error opportunities. The minimum error opportunities computed is 18 and the maximum is 30.

3.3.3 "No interruption"

Error count

Any action on the non-PC steps not following the correct sequence was counted as an error. These error counts include errors from trials with no interruptions, i.e. Nil, and also errors before an interruption in trials with an interruption.

Error opportunity

The error opportunities for each participant might be different due to trial variations in interruption position. Each participant could have a minimum of 181 error opportunities or a maximum of 193.

Table 6.5 shows the number of non-PCEs and their respective mean error rates according to the three interruption position categories. About half of the errors in the "no interruption" category (117 out of 228) occurred at "Selector DP" (step [2]), however, the interruption manipulation does not involve an interruption just before the step. Therefore, errors at "Selector DP" are excluded from the following analysis and their occurrence will be addressed later in the chapter.

	Interruption position			Total no. of non-PCE	
	Immediately after	Later after	No interruption		No interruption*
Total no. of non-PCE (Total no. of opportunities)	31 (124)	13 (750)	228 (5605)	272	111 (5264)
Mean error rate (SD)	25% (25)	1.7% (2.6)	4.1% (2.2)		2.1% (1.4)

** This "no interruption" category excludes errors at "Selector DP" step*

Table 6.5: Number of non-PCEs and their mean error rates with respect to the different interruption positions.

A one-way repeated ANOVA on the error rates yielded a large significant main effect of interruption position. The Greenhouse-Geisser correction was used because of violation of sphericity ($F(1.015, 30.439) = 25.6$, $p < .001$; Eta squared = .46). Post hoc comparisons with Bonferroni correction showed a reliable difference between "immediately after" and "later after", the difference between "immediately after" versus "no interruption" was also significant. No reliable difference was detected between "later after" and "no interruption". The result suggests that interruption position had the same disruptive effect on other non-PCEs as on PCEs: an interruption occurring just before a task step was more likely to result in an error than an earlier interruption or no interruption at all.

3.4 PCE and non-PCE resumption pattern

Given that the above results suggest that interruption position had the same general disruptive effect on PCE as well as non-PCE, are there any characteristics suggesting PCE might be different to non-PCE? This question is best addressed by looking at the pattern of task step resumption after an interruption. All 37 PCEs that occurred immediately after an interruption had the same pattern of resumption; that they were errors omitting the PC step and moving on to the Call Centre task. In contrast there was a qualitatively different resumption pattern among the non-PCEs that occurred immediately after an interruption; 45% (14 out of 31) involved resuming to a major task step just before the interruption (the previously operated compartment), and the remaining lacked a consistent pattern of resumption. The difference in resumption pattern between PCE and non-PCE suggests cues in different task steps play a role in aiding task resumption; the presence of a false completion signal at the PC step lures one to carry out the next task, but such a cue was not present at the non-PC steps in the task.

3.5 Resumption times between different task steps

The time it took one to resume the primary task after an interruption was analysed. Resumption times were taken only for *correct* task resumptions. For the same reason as in the previous experiment, resumption times for incorrect performance might not accurately reflect a participant genuinely trying to remember what he/she was about to do. As a consequence, these resumption times for incorrect performance were not included in calculating the mean resumption times.

The purpose of this analysis is to examine whether there was a difference in resumption times between the PC step and non-PC steps. Like the analysis of the previous experiment, mean task resumption times were calculated for each participant according to the PC step and non-PC steps. Because only correct task resumption data were included, the mean resumption times calculated for each participant were not necessarily an average of 4 (4 trials for each of the two interruption positions) but might be less in some cases if there were incorrect resumption times. Since there were three participants who did not resume the PC step after interruption Z for all 4 trials, no resumption times were obtained for the PC step. Data from these three participants were excluded from the analysis since comparison of mean resumption times was not possible between PC step and non-PC steps. As a consequence, the analysis is based on data from 28 participants who resumed correctly at least once for the PC step and the non-PC steps.

Table 6.6 below shows the mean resumption times for the PC step (after interruption Z) and non-PC steps (after interruption Other). A reliable difference was detected between position Z and position Other, $t(27) = 7.179, p < .001$.

	PC step	Non-PC step
Mean resumption time	3.4s	7.4s
(SD)	(1.5)	(2.9)

Table 6.6: Mean resumption times of PC step and non-PC step.

The result further supports the notion of usage of cues in the task environment upon resumption. Although the presence of a false completion was more likely to lure one in omitting the PC step, when one did not make the error the signal results in faster resumption by indicating where one was in the task sequence. Such a cue was not present after an interruption at other non-PC steps in the task.

However, there were differences in terms of the location and button sizes between the "Clean" button and the various "Selector" buttons. These differences might contribute to the difference found in the resumption times due to different physical movement times. As a consequence, Fitts' Law was used to calculate the approximate physical movement times for resuming the PC step and the non-PC steps (any of the "Selector" steps) after an interruption. This is meant to be a rough approximation, so the distance of the movement measured is based on the assumption that the starting position is the location of the "Pack" button of the interrupting Doughnut Packing task.

Fitts' Law calculation yielded 0.31 seconds and 0.67 seconds for the movement times from the "Pack" button to the "Clean" button and "Selector" buttons respectively (see Appendix E for details of the calculations). These movement times were taken out from the resumption times for each participant. Table 6.7 shows the resumption times with the estimated movement times taken out.

	Interruption position	
	Z	Other
Mean resumption time	3.0s	6.7s
(SD)	(1.5)	(2.9)

Table 6.7: Mean resumption times (minus estimated movement times) after interruption Z and Other.

A reliable difference was still present between the resumption times, $t(27) = 6.536$, $p < .001$.

3.6 Categorisation of non-PCE

Since there were a substantial number of other non-PCEs it is necessary to subject these errors to finer grain analysis. There were a total of 274 errors at the non-PC steps in the task. About 77% (212 out of 274) occurred at the one of the five "Selector" steps and the remaining 23% (62 out of 274) occurred at various steps in the task.

An examination of the error patterns based on the first mouse-click at a task step identified a number of error kinds, which are described as follow:

3.6.1 Skip-selector error

Error involving an action on the correct compartment in the task sequence, but omitting the corresponding "Selector" step to activate the compartment. This kind of error can be described as the correct task sequence execution but incorrect usage of the device interface.

3.6.2 Incorrect-sequence error

Error involving an action on the incorrect compartment in task sequence, but correct action at the corresponding "Selector" step to activate the compartment. This kind of error can be described in terms of incorrect task sequence execution but correct usage of the device interface.

3.6.3 Skip-and-incorrect error

Error involving an action on the incorrect compartment *and* omitting the corresponding "Selector" step to activate the compartment. This error kind can de described as incorrect task sequence execution and incorrect usage of the device interface.

3.6.4 Miscellaneous error

These errors occurred at various different steps in the task sequence with no particularly meaningful description in terms of the current study. Examples of these error include proceeding to the next task step while waiting for a progress bar to fill up, and incomplete specification of parameters in one of the compartments.

Table 6.8 shows the relative proportions of the four identified error kinds.

	Selector-step errors			Miscellaneous errors
	Skip-selector error	Incorrect-sequence error	Skip-and-incorrect error	
No. of errors	172	33	7	62
Proportion	63%	12%	2.5%	22.5%

Table 6.8: The relative proportion (%) of the different identified error kinds.

3.7 Interruption position effect on selector-step errors

The following analyses explore the effect of interruption position on selector-step errors, which made up the majority of the non-PCEs. The purpose of the analysis is to examine the errors in finer detail, in order to see if they behave the same or differently to PCEs in relation to interruption position.

3.7.1 Skip-selector errors

The first kind of error to examine is skip-selector error. Among the 172 errors approximately 67% (115 out of 172) were skip-selector errors occurring at the first compartment — Dough Port ("Selector DP"). Since manipulation of the interruption did not involve a position occurring before the Dough Port compartment, these skip-selector errors were excluded from the analysis. The remaining 57 errors were categorised into "immediately after interruption", "later after interruption" and "no interruption". The logic of counting the errors and their opportunities is the same as for all non-PCEs earlier, the only difference is that the counting was carried out on the respective "Selector" steps only.

Table 6.9 below shows the number and the mean error rates for skip-selector errors in relation to the different interruption positions. An error rate for each interruption position category (no. of errors / no. of opportunities) was computed for each participant. On a descriptive level, the trend of the data suggests that interruption did not increase the error as it did on PCE. A one-way repeated ANOVA yielded no significant effect of interruption position, with Greenhouse-Geisser correction, $F(1.676, 50.265) = 0.814$, $p = .429$.

	Interruption position		
	Immediately after	Later after	No interruption
Total no. of Skip-selector error (Total no. of opportunities)	3 (99)	9 (151)	45 (1114)
Mean error rate (SD)	3.2% (10.0)	6.0% (12.7)	4.0% (5.0)

Table 6.9: Number and mean error rates of skip-selector errors according to different interruption position.

3.7.2 Incorrect-sequence errors and PCEs

The second kind of error to examine is the incorrect-sequence error. The skip-and-incorrect errors were also included in the analysis since they share a common characteristic with the incorrect-sequence error, namely, selecting the incorrect step in the task sequence. These errors were grouped into the three interruption position categories: "immediately after interruption", "later after interruption" and "no interruption". The logic of counting the occurrences of the error and calculating the error opportunities are the same as the skip-selector errors, except that the counting also include the "Process" step, which was sometimes preceded by an interruption (position T), as well as the respective "Selector" steps.

There were a total of 40 incorrect-sequence errors (including 7 skip-and-incorrect errors). 2 of them occurred at the Dough Port compartment and they were excluded from the analysis since they could not be preceded by an interruption. Table 6.10 shows the number and mean error rates for incorrect-sequence errors and PCE in relation to different interruption positions. The error rate of incorrect-sequence error for each participant was calculated according to the respective opportunities.

Error rates of incorrect-sequence error and PCE were subjected to a 2 × 3 (error type × interruption position) repeated ANOVA yielding a significant main effect of interruption position, $F(1.253, 37.582) = 30.903$, $p < .001$, with Greenhouse-Geisser correction. The main effect of error type was also significant, $F(1, 30) = 5.391$, $p = .027$. The interaction was found not to be significant, $F(1.259, 37.756) = .007$, $p = .956$, with Greenhouse-Geisser correction.

	Interruption position		
	Immediately after	Later after	No interruption
Total no. of Incorrect-sequence error (Total no. of opportunities)	28 (124)	2 (250)	8 (1331)
Mean error rate (SD)	22.6% (22.7)	1.1% (4.2)	0.6% (1.8)
Total no. of PCE* (Total no. of opportunities)*	37 (124)	11 (124)	8 (93)
Mean error rate* (SD)*	29.8% (32.5)	8.9% (19.9)	8.6% (21.0)

**These figures are the same figures from Table 6.4.*

Table 6.10: The number and mean error rates of incorrect-sequence error and PCE according to the different interruption positions.

A simple effect analysis on incorrect-sequence error yielded a significant effect of interruption position, $F(1.06, 31.789) = 26.951$, $p < .001$. Post hoc comparisons with Bonferroni corrections revealed a reliable difference between "immediately after" versus "later after" and "immediately after" versus "no interruption", but no reliable difference between "later after" versus "no interruption". The results suggest that interruption position has an effect on the occurrences of incorrect-sequence error; furthermore, the non-significant interaction suggests that there was no difference between PCE and incorrect-sequence error with regards to the interruption effect.

3.8 Errors at the Dough Port selector step

A total of 117 errors occurred at the very first compartment step in the task sequence, namely, the Dough Port (DP) compartment. Almost all of the errors (98%; 115 out of 117 errors) were identified as skip-selector errors while the remaining two were classified as incorrect-sequence. The pattern of the data suggests that participants mostly carried out the task sequence correctly at the DP step but kept using the device interface incorrectly. The occurrence of this error is discussed in more detail later in the chapter.

4 Discussion

The obtained overall error rate of 16% for PCEs (56 out of 341 opportunities) in the current experiment exceeds the 5% systematicity criterion and it is considerably higher than that generated in Experiment 6a (about 1%; 3 out of 220 opportunities). The current error rate obtained from trials without interruptions (8.6%; 8 out of 93 opportunities) is also comparable to the 9.3% (13 out of 140 opportunities) obtained in Byrne & Bovair's (1997) Experiment 1, which did not have working memory load manipulation. We can claim with reasonable confidence that the current paradigm has successfully generated PCEs at a level that allows investigation in a laboratory setting.

A significant difference was found in the PCE rate between the different interruption positions. As predicted in the original hypothesis, the results suggest that it is more likely for a PCE to occur when the task was being interrupted *just-before* (position Z) the execution of the PC step than any other positions during the task (positions P, Q, R, S or T). This finding can be explained in terms of Altmann & Trafton's framework of AGM; the low PCE rate in trials with no interruptions or with interruptions at other earlier positions in the task is because upon task resumption there are still remaining procedures to be carried out until task completion, and when the preceding step of the PC step is executed this provides associative priming to the PC action. The results suggest that as long as the associative link between the PC step and its preceding step are not disrupted, interruptions occurring early in the task are not likely to increase the occurrences of PCE. Evidence comes from the data showing that there was no reliable difference in trials with interruption Other and no interruptions in terms of PCE occurrences; this finding is consistent with the AGM model's notion of associative priming between task steps in routine procedural tasks.

On the other hand, interruptions just before the PC step will have disrupted the associative priming from its preceding step and upon task resumption the goal of executing the PC step will have decayed below retrieval threshold. Moreover, the task environment indicates the task has been completed with a signal prompting one to collect the next order from the Call Centre. As a consequence, when returning to the interrupted task, the goal of moving on to the next task is more likely to be primed by the completion signal than the goal of executing the PC step, because its associative links with the preceding step had been disrupted and underwent considerable decay. This highlights a situation where one relies on cues in the environment to resume a disrupted task; however, the environmental cues are priming a new task goal, whereas the PC goal requires priming from memory cues. This provides support to the idea that the execution of a PC step requires some kind of "deliberate" process so that the error can be avoided most of the time (Altmann & Trafton, 2002; Young, 1994).

In the previous experiment, the difference in resumption times between the PC step and non-PC steps was found to be significant in the condition with the 75-sec mental arithmetic interruption. The same pattern of results was found: resumption of the PC step is quicker than resumption of the non-PC steps. This provides further support for the explanation that when resuming after an interruption at any of the positions *during* the task (position P, Q, R, S or T), the encoding demands for these sub-task points were greater because there was no differential information indicating the current state of the task. On the other hand, the interruption at position Z occurred after the task was completed and the encoding demand for the remaining PC step should be relatively less because it is a 'standalone' step with no similar steps to be differentiated from. However, this finding might be a by-product of the design of the experimental task rather than general to all procedural PC tasks. For example, in the general case of a photocopying task, completion of other task steps (e.g. entering how many copies are to be photocopied) might also result in cues (e.g. feedback from a small screen) differentiating one's current state in the overall task, which could facilitate one's resumption time after an interruption. Therefore, in this photocopying example, it is not just the PC step that has differential cues but other task steps as well.

Nonetheless, the main message from the observed faster resumption at the PC step than at non-PC steps in the current experiment is that it suggests the participants use cues in the task environment upon resuming an interrupted task. This is also consistent with the data on the qualitative pattern of resumption after interruption position Z: that all occurrences of forgetting to execute the PC step involved moving on to the Call Centre rather than carrying out any other actions. This suggests the presence of the completion signal has achieved its purpose in urging one to move on to the next task goal.

Further analysis into other procedural errors in the task has shed some light onto the question regarding whether PCEs are unique to the effect of interruption position. When treating all other errors as non-PCEs the result suggests that they were sensitive to the position of the interruption; more errors occurred immediately after an interruption. The same general disruptive effect of interruption was obtained for non-PCEs as well as for PCEs.

When the non-PCEs were subjected to finer grain analysis categorising them into different error types by their pattern of occurrences, it was found that incorrect-sequence errors were also sensitive to interruption position. Furthermore, the interaction between error type and interruption position was found not to be significant suggesting that PCEs and incorrect-sequence errors were equally sensitive to the disruptive effect of interruption. Both kinds of error were more likely to occur immediately after an interruption than later after or having no interruption at all. This suggests that resuming the PC step immediately after an interruption is similar to resuming the correct sequential step in the task; in other words, the correct resumption depends on whether the associative links between consecutive task steps in memory remain intact or not. The similarity between PCE and incorrect-sequence error with respect to interruption position suggests that, immediately after an interruption, the associative link between the to-be-carried out step and its preceding step is disrupted. As a result, one forgets which task step had been executed which leads to difficulty in retrieving the next correct step.

In contrast to PCE and incorrect-sequence error, the occurrence of skip-selector error was found to be independent of the effect of interruption position. The error was not more likely to occur immediately after an interruption. This kind of error has an interesting property that it is not an error of executing the incorrect task sequence, but an error of using the interface incorrectly. Strong evidence supporting this notion comes from data showing that almost all of the errors at the very first compartment ("Selector DP") were skip-selector errors and only a few were incorrect-sequence errors. This suggests that one remembers quite well the Dough Port compartment as the first operation in the task sequence, but still forgets the correct way of using the interface: selecting the desired compartment using the "Selector" first. The occurrence of the skip-selector error provides a neat contrast to PCE highlighting that PCE is not a result of deficiency in one's knowledge about the interface, but an error of forgetting one's place in the task sequence.

Experiment 6c[7]: The effect of interruption duration

1 Introduction

Experiment 6b successfully generates PCE and produces the interruption position effect on the error as predicted by the AGM model. The disruptive effect of an interruption is explained in terms of the decay of activation of goal memory. The time-based decay process specified in the AGM model suggests that the longer a memory item is left unrehearsed (or unretrieved), the more its activation decays. The next experiment in the series investigates the effect of the duration of an interrupting activity on PCE.

This experiment is a continuation of Experiment 6b examining the effect of two different interruption durations: 45 seconds and 15 seconds. Based on the AGM model, the following hypothesis is made about the effects of interruption position and duration:

With an interruption duration long enough for a memory item to decay below retrieval threshold, the interruption position effect should persist — there is more likely to be a PCE immediately after an interruption just before the PC step than an interruption occurring earlier in the task. On the other hand, the effect should disappear if the interruption duration is too short for substantial decay to take place — interruptions occurring just before the PC step should not be any more likely to give rise to PCEs than interruptions occurring earlier in a task. More specifically, we should expect an interaction effect between interruption position and duration on PCE if a 15-sec interruption is too short for any substantial decay of memory.

Furthermore, analysis on non-PCEs from the previous experiment suggests that some of these errors are sensitive to the disruptive effect of interruption. Therefore, it is expected that the same general result for non-PCEs will be obtained in the current experiment. More specifically, the occurrences of skip-selector errors should be independent of interruption. The occurrences of incorrect-sequence errors should be sensitive to the effect of interruption such that we should expect an interaction effect between

[7] This experiment was administered by Aliza Abeles as part of her undergraduate project. The design and data analysis of the experiment are the author's original work.

interruption position and duration if the 15-sec interruption is too short for memory of the next task step to decay.

2 Method

2.1 Design

The experiment is a mixed design. It has two independent variables; the within-subject variable is interruption position with three levels; Z, Other and Nil. The between-subject variable is the duration of the interruption with two levels; Short (15 seconds) or Long (45 seconds).

2.2 Participants

There were 24 participants and they were randomly allocated to either condition Short or Long (12 in each condition). Participants were university undergraduates and postgraduates, age ranging from 20-24 with a mean age of 21.4 years. None of the participants had taken part in the previous experiments.

2.3 Materials and procedures

Adjustments were made to the computer program to change the duration of the interruption to 15 seconds for the Short interruption condition and 45 seconds for the Long interruption condition. In all other ways, the tasks, apparatus and procedures were as described in Experiment 6b.

2.4 Measures

The dependent measure of primary interest is the same as Experiment 6b, namely, the number of PCEs. Resumption times and other errors are also of interest and recorded.

3 Results

The data from one participant in the Long condition and two participants in the Short condition were removed as they made a PCE on every trial suggesting that they had not correctly understood the task.

3.1 Overall errors

A total of 286 errors were obtained across the 21 participants. More than half of the participants (16 out of 21) committed at least one PCE yielding a total of 71 PCEs. About 25% of the total errors were accounted for by the occurrences of PCE. A Binomial test (p = 0.05) shows the occurrences of PCE is above chance level, z = 15.38, p < .001.

As in Experiment 6b, there were 11 trials for each participant and this gives 231 error opportunities for each task step in calculating the error rates. Figure 6.11 below shows the distribution of the error rates for each task step.

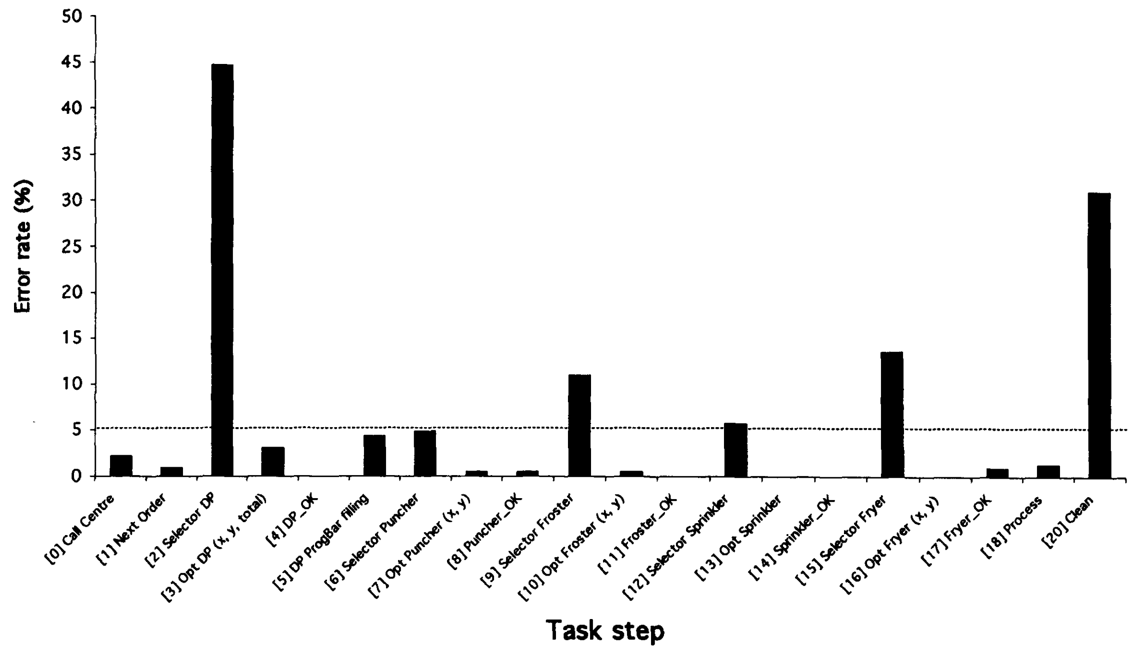


Figure 6.11: Error rate of each task step.

A consistent pattern of error rates across the task steps as Experiment 6b was obtained; the "Selector DP" step has the highest error rate (about 45%) followed by the PC step ("Clean"; about 30%). Error rates at the other "Selector" steps either exceed the 5% systematicity criterion or just about reach the criterion. Errors at the "DP ProgBar filling" are also near the criterion but they are not to be further pursued for the same reason as mentioned in Experiment 6b; the occurrence of these errors suggests the participants may be impatient. Further pursuit of this error is not the focus of this study and is probably not going to help the understanding of PCEs.

3.2 Interruption position effect on PCE

Table 6.11 shows the number of PCEs and their error rates in relation to the different interruption positions and durations.

		Interruption position		
		Z	Other	Nil
Short interruption (15secs)	No of PCEs (no. of opportunities)	16 (40)	8 (40)	6 (30)
	Mean error rate (SD)	40.0% (39.4)	20.0% (30.7)	20.0% (28.1)
Long interruption (45secs)	No of PCEs (no. of opportunities)	23 (44)	12 (44)	6 (33)
	Mean error rate (SD)	52.2% (48.0)	27.3% (30.5)	18.2% (31.1)

Table 6.11: Number of PCEs and their error rates with respect to different interruption positions in the two experimental conditions.

PCE rates were calculated for the three different interruption positions for each participant in the two conditions. The error rates were analysed using a mixed 2 × 3 ANOVA. There was a significant main effect of interruption position ($F(1.504, 28.568) = 7.166$, $p = .002$) with a large effect size (Eta squared = .274). Greenhouse-Geisser correction was used due to violation of sphericity. Planned contrasts, across the two conditions, revealed significant difference between Z and Other ($t(20) = 2.576$, $p = .018$) and Z versus Nil ($t(20) = 3.324$, $p = .003$). There was no significant difference between Other and Nil ($t(20) = .946$, $p = .355$).

There was no main effect of interruption duration ($F(1,19) = .215$, $p = .648$). The data in both duration conditions, show that in trials where there was no interruption, or where the interruption occurs in any of the Other positions, one would not expect any difference between the groups in terms of the error rate. In trials where the interruption occurs immediately before the PC step, there was an increase in the error rate. Although the predicted effect of interruption duration on PCE was not significant, the trend of the data shows a relatively higher PCE rate in each interruption position in condition Long than condition Short.

The interaction, interruption position × duration, was not statistically significant ($F(1.504, 28.568) = .436$, $p = .595$) suggesting that the effect of interruption position on PCE did not differ between the two interruption durations.

3.3 PCE and non-PCE resumption pattern

Two of the 39 PCEs that occurred after interruption Z had slightly different resumption patterns to the operationalised criterion: the "Next Order" button was clicked before the Call Centre task was carried out. Despite the difference, they show a qualitatively consistent resumption pattern: that of getting the next order rather than performing other actions in the task. This suggests a systematic bias towards committing a PCE after interruption Z. The systematic bias suggests that cues in the external task environment might prompt participants about the completion of the task. In contrast, over 50% (15 out of 27) of the non-PCEs that occurred immediately after an interruption involved repeating the preceding task step, i.e. selecting a previously operated compartment. There was not a consistent resumption pattern in the remainder. The trend of resumption patterns in the current experiment is consistent with that of Experiment 6b.

3.4 Resumption times between different task steps

Like the analyses on resumption times carried out in Experiment 6b, the current analyses were performed on correct resumption data only to eliminate potential biases from incorrect resumption behaviour. The exclusion of reaction times from incorrect resumptions yielded valid data from 8 participants in condition Short and 6 participants in condition Long. These participants resume correctly at least once for both PC step and non-PC steps.

Table 6.12 shows the mean resumption times for PC and non-PC steps in the Short and Long conditions. A 2 × 2 mixed ANOVA yielded a large significant main effect of task step on resumption times, $F(1, 12) = 33.904$, $p < .001$, Eta squared = .739. There was no significant difference between conditions Short and Long, $F(1, 12) = 1.619, p = .227$. The interaction, interruption position × duration, was found not to be significant, $F(1, 12) = .001, p = .973$.

	PC step	Non-PC step
Short	2.8s (0.6)	5.7s (1.8)
Long	3.8s (1.1)	6.6s (2.7)

Table 6.12: Mean resumption times (standard deviation in bracket) after interruptions Z and Other.

However, in order to avoid possible confounding effects from different physical movement times between the interruption positions, the same movement times estimated in Experiment 6b (0.31s and 0.67s for PC step an non-PC step respectively) were removed from the resumption times. Table 6.13 shows the mean resumption times with the estimated movement times removed.

	PC step	Non-PC step
Short	2.5s (0.6)	5.0s (1.8)
Long	3.5s (1.1)	6.0s (2.7)

Table 6.13: Mean resumption times with estimated movement times removed (standard deviation in bracket) after interruption Z and Other.

The same results were obtained: a large significant main effect of interruption position was found ($F(1, 12) = 25.93$, $p < .001$, Eta squared = .684); there was no reliable difference between conditions Short and Long ($F(1, 12) = 1.619$, $p = .227$); and the interaction was found not to be significant ($F(1, 12) = .001$, $p = .973$). Simple effect analysis yielded a reliable difference in conditions Short ($t(7) = 5.044$, $p = .001$) and Long ($t(5) = 2.681$, $p = .044$).

3.5 Interruption position effect on non-PCE

Just as Experiment 6b, the effect of interruption position was also assessed against the obtained non-PCEs. There were a total of 215 non-PCEs across the two interruption duration conditions: 84 in condition Short and 131 in condition Long. For each condition, the non-PCEs were categorised into "immediately after interruption", "later after interruption" and "no interruption" (including errors *before* an interruption as well as errors from trials with no interruption). The procedures for counting the error occurrences and calculating their respective error opportunities are the same as Experiment 6b.

Table 6.14 below shows the number and mean error rates in relation to the different interruption position. In the two conditions, more than half of the errors in the "no interruption" category (41 out of 67 in condition Short, 62 out of 107 in condition Long) consist of errors occurring at "Selector DP" (step [2]); for the same reason as the previous experiment's analysis, these errors and their error opportunities are excluded from the following analysis.

		Interruption position			Total no. of non-PCE	
		Immediately after	Later after	No interruption		No interruption*
Short	Total no. of non-PCE (no. of opportunities)	10 (40)	7 267	67 (1783)	84	26 (1673)
	Mean error rate (SD)	25% (23.6)	2.6% (4.0)	3.8% (2.0)		1.6% (1.4)
Long	Total no. of non-PCE (no. of opportunities)	18 (44)	6 (261)	107 (1994)	131	45 (1873)
	Mean error rate (SD)	41% (30.2)	2.4% (3.2)	5.4% (2.6)		2.4 (1.5)

**This "no interruption" category excludes errors at "Selector DP" step*

Table 6.14: Number of non-PCEs and their mean error rates with respect to the different interruption positions.

A 2 × 3 mixed ANOVA was performed on the mean error rates and yielded a large significant main effect of interruption position, $F(1.016, 19.302) = 25.895$, $p < .001$, Eta squared = .577. The assumption of sphericity was violated; therefore, the Greenhouse-Geisser correction was adopted. Post hoc comparisons with Bonferroni correction showed a reliable difference between "immediately after" and "later after", and "immediately after" vs "no interruption". No significant difference was detected between "later after" and "no interruption".

The main effect of interruption duration was found not to be significant, $F(1, 19) = 1.925$, $p = .181$. There was no significant interaction between interruption position and duration, $F(1.016, 19.302) = 1.675$, $p = .211$.

3.6 Categorisation of non-PCE

The obtained non-PCEs in the two experimental conditions were classified into different kinds of errors as in Experiment 6b: the identification of the errors was based on the first mouse-click at a task step. The same four categories of errors were identified (skip-selector, incorrect-sequence, skip-and-incorrect and miscellaneous) as the previous study; Table 6.15 summarises the number of occurrences and the approximate relative proportions in condition Short and Long. Consistent with the pattern of results in Experiment 6b, the majority of the non-PCEs were selector-step errors. The relatively small proportion of miscellaneous errors consists of errors occurring at various task steps.

		Selector-step errors			Miscellaneous errors
		Skip-selector error	Incorrect-sequence error	Skip-and-incorrect error	
Short	No. of errors	61	4	5	14
	Proportion	72%	5%	6%	17%
Long	No. of errors	91	17	1	22
	Proportion	69%	13%	1%	17%

Table 6.15: The number and relative proportion (%) of the different error kinds in condition Short and Long.

3.7 Interruption position effect on selector-step errors

As in the analysis carried out in Experiment 6b, the obtained selector-step errors are subjected to finer detail analysis according to the identified error categories. The same logic and procedures of error count and error opportunities calculations are adopted.

3.7.1 Skip-selector errors

In both interruption duration conditions, more than half of the skip-selector errors occurred at the “Selector DP” step: 40 out of 61 in condition Short, and 61 out of 91 in condition Long. As in the way these errors were dealt with in Experiment 6b, they are excluded from the following analysis because the interruption manipulation did not occur before the “Selector DP” step. Analysis including these errors would confound comparisons between the different interruption positions.

Table 6.16 shows the number and mean error rate of skip-selector errors with respect to the interruption positions.

		Interruption position		
		Immediately after	Later after	No interruption
Short	Total no. of Skip-selector error (Total no. of opportunities)	3 (33)	4 (56)	14 (351)
	Mean error rate (SD)	9.2% (14.9)	7.0% (12.0)	4.0% (4.7)
Long	Total no. of Skip-selector error (Total no. of opportunities)	4 (36)	5 (51)	21 (397)
	Mean error rate (SD)	9.8% (17.8)	10.9% (17.3)	5.4% (5.8)

Table 6.16: Number and mean error rates of skip-selector errors according to different interruption position in condition Short and Long.

A 2 × 3 mixed ANOVA was carried out on the mean error rates. The main effect of interruption position ($F(1.348, 25.608) = .84$, $p = .401$, with Greenhouse-Geisser correction) and interruption duration ($F(1, 19) = .35$, $p = .561$) were not significant. There was no significant interaction effect, $F(1.348, 25.608) = .087$, $p = .842$. The results suggest that the occurrences of skip-selector errors were not sensitive to the effect of interruption position in condition Short and Long.

3.7.2 Incorrect-sequence errors

There was one incorrect-sequence error occurring at "Selector DP" step in condition Short, and it was excluded from the following analysis because it was not preceded by the interruption manipulation. None of the incorrect-sequence errors in condition Long occurred at "Selector DP" step.

As in the analysis in Experiment 6b, skip-and-incorrect errors were included in the analysis of incorrect-sequence errors. This is because both kinds of error have a commonality of executing the incorrect sequential step in the task. This yielded a total of 8 and 18 incorrect-sequence errors in condition Short and Long respectively. Table 6.17 shows the number and mean error rates of incorrect-sequence error in the two experimental conditions with respect to the interruption positions.

		Interruption position		
		Immediately after	Later after	No interruption
Short	Total no. of Incorrect-sequence error (Total no. of opportunities)	3 (40)	1 (89)	4 (421)
	Mean error rate (SD)	7.5% (12.1)	1.3% (4.0)	0.9% (2.9)
Long	Total no. of Incorrect-sequence error (Total no. of opportunities)	13 (44)	0 (87)	5 (474)
	Mean error rate (SD)	29.5% (29.2)	0% (0)	1.1% (1.2)

Table 6.17: The number and mean error rates of incorrect-sequence error according to the different interruption positions in the two experimental conditions.

A 2 × 3 ANOVA on the mean error rates yielded a large significant main effect of interruption position, $F(1.003, 19.054) = 12.562$, $p = .002$, Eta squared = .398, with Greenhouse-Geisser correction. There was a marginal significant main effect of interruption duration, $F(1, 19) = 4.227$, $p = .054$. The interaction, interruption position × duration, was also found to be significant, $F(1.003, 19.054) = 5.127$, $p = .035$, Eta squared = .212, with Greenhouse-Geisser correction. The significant interaction suggests that the effect of interruption position on incorrect-sequence errors might be sensitive to the effect of interruption duration.

Simple effect analysis revealed no significant effect of interruption position in condition Short, $F(1.007, 9.062) = 2.33$, $p = .161$, with Greenhouse-Geisser correction. A nonparametric test was used in analysing the effect of interruption position in condition Long because of the occurrence of zero variance in one of the interruption position cells.

Post hoc comparisons using Wilcoxon Signed Rank Test, with Bonferroni correction, yielded reliable differences between " "immediately after" versus "no interruption".

The results suggest that there was an effect of interruption position on incorrect-sequence error in condition Long but not in condition Short. Moreover, the interruption position effect on the error was found to be consistent with the predicted direction: an increase in the error immediately after an interruption. The marginally significant difference between "later after" and "no interruption" is contrary to the predicted trend. However, interpretation of this result needs caution because the difference between the two interruption positions is relatively small, and the floor rate of error in "later after" is likely to be contributing most to the significant difference.

3.8 Errors at "Selector DP" step

Across the two interruption duration conditions, a total of 103 errors occurred at the very first compartment step — "Selector DP". The same consistent pattern as Experiment 6b was obtained; that over 95% (101/103) of the errors at "Selector DP" step were skip-selector errors and the remaining two were incorrect-sequence error and miscellaneous error.

4 Discussion

The results suggest that interruptions occurring just before the PC step are more likely to result in PCEs than interruptions occurring at other positions in the task. The effect is robust in that it is replicated in the current experiment. Both the 15-second and 45-second interruptions are of sufficient length to result in an increase in PCEs when interrupted at position Z – almost twice the rate – than at Other and Nil. Although the results suggest a trend in the predicted direction, that the length of the interruption itself can also influence the error rate: higher in the 45-second duration than the 15-second duration, the results did not show the anticipated interaction effect between interruption position and duration. This suggests that the shorter interruption is not short enough to have a significantly reduced effect on the PCE rate.

Although the trend of the resumption time data suggests faster resumption times after the 15-second interruption than the 45-second interruption, there was no statistical difference in resumption times between the two interruption durations. Monk & Trafton (2004) showed that even very brief interruption of 5 seconds results in reliably slower resumption time than 1/4-second and 1-second interruptions, and the current findings are consistent with the general observation in that cognitive overheads, in terms of increased resumption times, are associated with interruptions. However, conclusion regarding the difference in resumption times between the 15-second and 45-second interruptions needs caution. This is because of the overall small sample of valid data for a between-subject comparison (six valid data points in condition Short and eight in condition Long), due to the exclusion of resumption times from incorrect resumption data in the current data set.

Faster resumption times of the PC step than other task steps are obtained and these are a replication of findings from previous experiments. The resumption patterns of PCEs and non-PCEs immediately after an interruption are also consistent with those obtained in Experiment 6b. These results provide support for the role of cue usage upon resuming an interrupted task.

Analysis on non-PCEs produced results consistent with those of Experiment 6b, suggesting that these errors are also sensitive to the disruptive effect of interruption. There were, generally, more errors immediately after an interruption than some time later after an interruption or no interruption at all. This was found to be the same for the 15-second interruption and the 45-second interruption. Finer grain analysis of non-PCEs obtained the same result on skip-selector error: the occurrence of this kind of error is not sensitive to the effect of interruption position. The same result was found in the two interruption duration conditions.

Consistent with the results from Experiment 6b, the effect of interruption position was found to be significant in the analysis of incorrect-sequence error. There were reliably more incorrect-sequence errors immediately after an interruption than some time later after an interruption or with no interruption at all. The difference in the occurrence of incorrect-sequence errors between the two different interruption durations was found to be marginally significant. Moreover, the significant interaction effect between interruption position and duration reveal that the interruption position effect was present in the 45-second condition but less so in the 15-second condition. This suggests that the disruptive effect of interruption on incorrect-sequence error has lessened with the shorter interruption duration. Although the trend of the data shows more errors immediately after an interruption than some time later after an interruption or no interruption at all, the increase in incorrect-sequence error rate with an immediately preceding interruption did not reach statistical significance. This finding is consistent with the prediction made from the AGM model: with short-enough interruption duration the to-be-resumed goal would have undergone the time-based decay process, but not so much that it is no longer the most active goal in memory. Therefore, the to-be-resumed goal still has a high-enough level of activation to be the most active memory element upon task resumption and gets retrieved successfully.

There was no significant difference in PCE rate between the two duration conditions in the current experiment: evidence that even relatively short interruptions can result in a significant increase in PCE rate immediately after an interruption. The initially predicted interaction effect between interruption position and duration was not obtained with respect to PCE; however, it was present with the analysis of incorrect-sequence error. Results from Experiment 6b suggest that both kinds of error — PCEs and incorrect-sequence errors — are sensitive to the interruption position effect. Current results confirm this finding but also further suggest that these two kinds of error are differentially sensitive to the duration of an interruption. The currently adopted short interruption duration of 15 seconds was found to be disruptive to incorrect-sequence errors but not PCEs. The observed difference between the two kinds of errors due to different interruption durations is a surprise finding and is not predicted by the AGM model. As a speculation, the observed difference between the two kinds of errors suggests that remembering to execute the PC step might be more vulnerable to interruption than remembering ones place in a task sequence during task execution. In terms of a time-based decay process, as specified in the AGM model, memory for a PC step might decay faster or have a lower activation value than memory for maintaining one's place in the task sequence. However, this explanation is highly speculative and requires further experimentation to test out the assumption before a conclusion can be made.

The current experiment obtained an overall PCE rate of about 34% (41 out of 121 opportunities) in the 45-second interruption condition and about 27% (30 out of 110 opportunities) in the 15-second condition. The PCEs rates in these two interruption durations are higher than that obtained in Experiment 6b's 75-second interruption, which is about 16% (56 out of 341 opportunities). The trend of the PCE rate in these different interruption durations suggests the contrary to the prediction made by the AGM model: the shorter an interruption duration the lesser the disruptive effect of the interruption, consequently, fewer PCEs obtained. The generally higher PCE rate in the current experiment might be explained by the fact that a different experimenter administered this experiment and that the participants were not paid for their participations. These factors might have had a motivational effect on the participants resulting in an overall higher PCE rate than the previous experiment. As a consequence, this makes comparison of the results of the current experiment and Experiment 6b impractical. All in all, results from the current experiment should be interpreted in their own right for this very reason.

In Experiment 6b the primary doughnut-making task was modified to address issues discussed in the analytic comparison made between the doughnut-making task and the Phaser task, such as the prominence of the PC step button, the separation of testing trials with an additional task, and the presence of a competing signal. The successful generation of PCEs using the modified version of the doughnut-making task allows investigation of interruption position (Experiment 6b) and interruption duration (Experiment 6c) to be carried out. However, the issue of a less-well-practised PC step identified in the analytical comparison remains to be addressed and this is what the next experiment set out to examine.

Experiment 6d: A change to the PC step

1 Introduction

In the Phaser task used in Byrne & Bovair's study, there is a peculiar feature to its task structure such that participants were required to repeat all the steps in the task, except the PC step, until the task is completed. We term this task structure a "cyclic" task structure because when the task is not completed, the participant could be stuck in a "loop" cycling through a set of procedures repeatedly. The PC step is executed *only* when the task has been completed, which means that the execution of the PC step is relatively less well practised than the other steps in the task. In contrast, the PC step in the doughnut-making task was to be executed after only one cycle of executing all the steps in the task, so the PC step receives as much practice as all other steps in the procedure. It is the very feature of a less-well-practised PC step that might be one of the contributing factors to the effectiveness of the Phaser task in provoking PCEs.

In the current experiment, the same primary tasks (Doughnut Machine and Call Centre), interruption task (Doughnut Packing Express) and interruption duration (75 seconds) as Experiment 6b were used. Only one modification was made to the Doughnut Machine task: the requirement of the PC action (cleaning the Doughnut Machine after each order) is *conditional* upon the presence of a notification signal at the beginning of each trial. The notification of the need to clean the Doughnut Machine only appears in 1/3 of the total number of trials; as a consequence, the need to carry out the PC step is less frequent relative to other steps in the task.

The objective of this experiment is to mimic the less-well-practised PC step in Byrne and Bovair's Phaser task by including a conditional PC step in the current task paradigm. It is expected that, with all else being equal, the effect of interruption position on PCE will be replicated and a general increase of PCE rate will be obtained relative to Experiment 6b. Moreover, a replication of the result on non-PCEs should also be expected: skip-selector error not affected by interruption, whereas the occurrence of incorrect-sequence error is sensitive to an immediately preceding interruption.

2 Method

2.1 Tasks

The tasks used in this experiment, namely the interrupting doughnut-packing task and the Call Centre task were identical to the ones used in Experiment 6b. The doughnut-making task used in the current experiment is a variant of the one used in the previous experiment. The basic procedures in operating the machine to make doughnuts are the same but the PC step (clicking the Clean button) is not required to be executed in every single trial; it is only required when specified at the beginning of 1/3 of the total number of trials. Figure 6.12 shows the screenshot at the beginning of a trial requiring the PC step to be carried out when the task is completed.

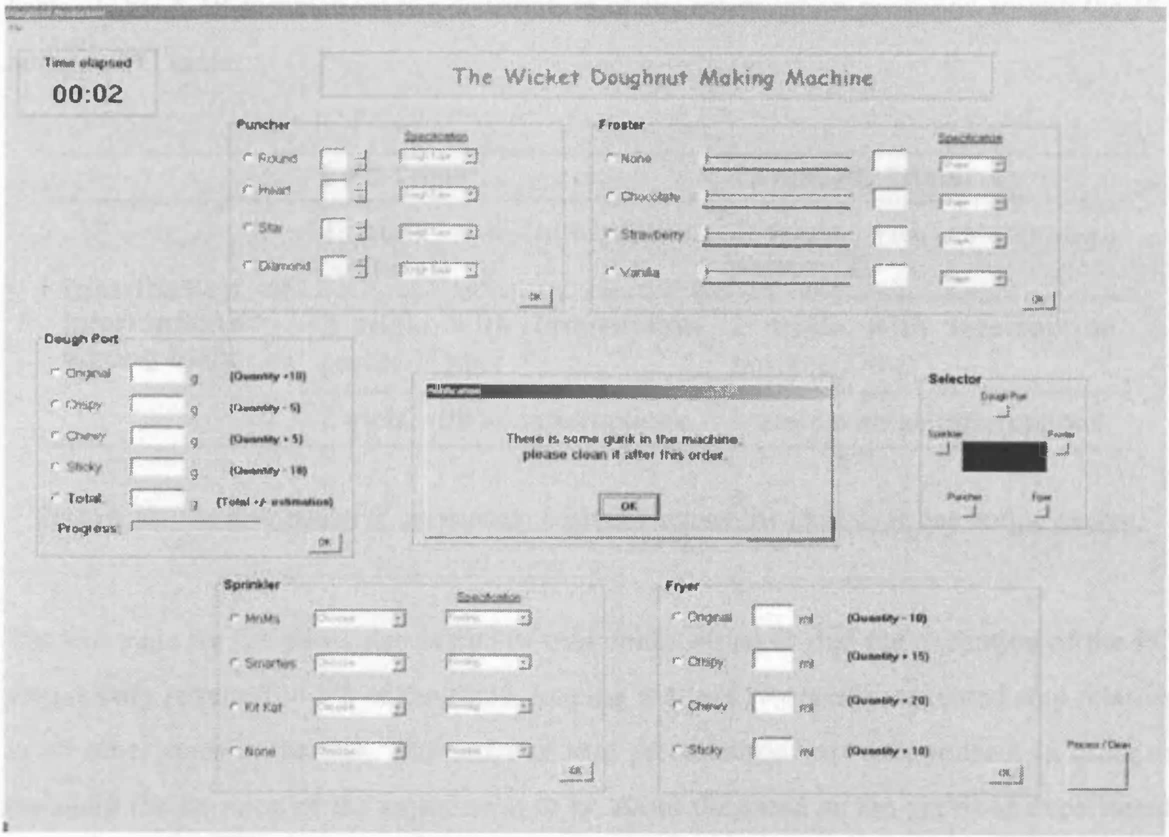


Figure 6.12: Screen transitions of the beginning of a PC trial.

In a trial requiring the PC step (a PC trial), when the Call centre task is completed and returning to the doughnut-making machine, once the "Show Order" button is clicked a message saying "*There is some gunk in the machine, please clean it after this order.*" will be displayed specifying that one should carry out the PC step of clicking the "Clean" button when the task is completed.

2.2 Design

The basic design of the current experiment is the same as Experiment 6b in that it is a within-subject design with one independent variable — interruption position — which has three levels: Just-Before (position Z), Other (position P, Q, R, S or T) and None (no interruption).

The current experimental design differs in terms of its trial composition. Each experimental session consists of 18 trials in the testing phase: 6 PC trials and 12 non-PC trials. Table 6.18 summarises the distribution of the interruption positions among the PC and non-PC trials.

	6 PC Trials	**12 non-PC Trials**
Distribution of interruptions among trials	2 trials with interruption position Z	2 trials with interruption position Z
	2 trials with interruption position Other	2 trials with interruption position Other
	2 trials with no interruptions	8 trials with no interruptions

Table 6.18: The distribution of interruption positions among the 18 trials in one testing session.

The rationale for the particular design of trial composition is that the execution of the PC step is only required in 1/3 of the trials, making it a less frequently executed step relative to all other steps in the task. The order of trial presentation was randomised. In order to maintain the duration of the experiment to be about the same as the previous experiment (about an hour), each trial only requires one to make one set of doughnut instead of two sets as in the previous two experiments.

2.3 Materials

The materials used were the same as Experiment 6b's, that is two different computer terminals: an HP Desktop and a Compaq desktop were used. The doughnut machine program and the doughnut-packing program were run on the HP desktop and the Call Centre was run on the Compaq desktop. The two computer terminals were arranged at a 90° configuration.

2.4 Participants

37 participants, either university undergraduate or postgraduate students, took part in this study. Ages ranged from 21 to 48 with a mean of 26.2. There were 18 female and 19 male. None of the participants had taken part in the previous experiments. All participants were paid £6 for their participations.

2.5 Procedure

The procedures were very similar to Experiment 6b with some slight changes. Participants read some documents describing the experiment then went through a demonstration and training phase. Participants first observed the experimenter performing both the doughnut-making and the doughnut-packing task separately. The experimenter performed two trials (one PC and one non-PC trial) on the doughnut-making task during the demonstration phase. Participants were then given three training trials on the doughnut-making task: two without and one with the interrupting doughnut-packing task. The two training trials without interruption were one non-PC and one PC trial; the trial with interruption was a PC trial. Any errors that occur in this training phase result in on-screen warning messages and beeps; participants were required to identify and correct the error in order to continue. The experimenter was present in the room with the participant during training, and the participant was allowed to ask questions about the tasks if necessary.

In the testing phase, participants were required to perform 18 trials in total and the experimenter left the participant to carry on the session alone at this point. The entire experiment took approximately an hour.

2.6 Measures

The same dependent measures as the previous experiments were used: number of PCEs, resumption times and other errors.

3 Results

A total of three participants' data was excluded from analysis. One participant did not perform the task according to the instructions, one participant terminated the experiment without completing the entire session and one participant had a lost data file.

The design of the current experiment consists of PC trials and non-PC trials; therefore, the results are presented and analysed according to the two trial types. Data from the PC trials are presented first, followed by data from the non-PC trials.

3.1 Analysis of PC trials

3.1.1 Overall errors

A total of 245 errors were obtained in the PC trials across the 34 participants. More than half of the participants (24 out of 34) made at least one PCE yielding a total of 56 PCEs. The PCEs account for about 23% of the total number of errors. A Binomial test shows the occurrence of PCE is above chance level, $z = 12.82$, $p < .001$. Figure 6.13 shows the error rate of each individual task step[8]. Since there were 6 PC trials for each participant, this gives 204 error opportunities for each task step in calculating the error rates.

[8] Task step [2] and [20] involve clicking on a modal dialog box, which do not afford errors to occur. Therefore, both steps are excluded from the analysis.

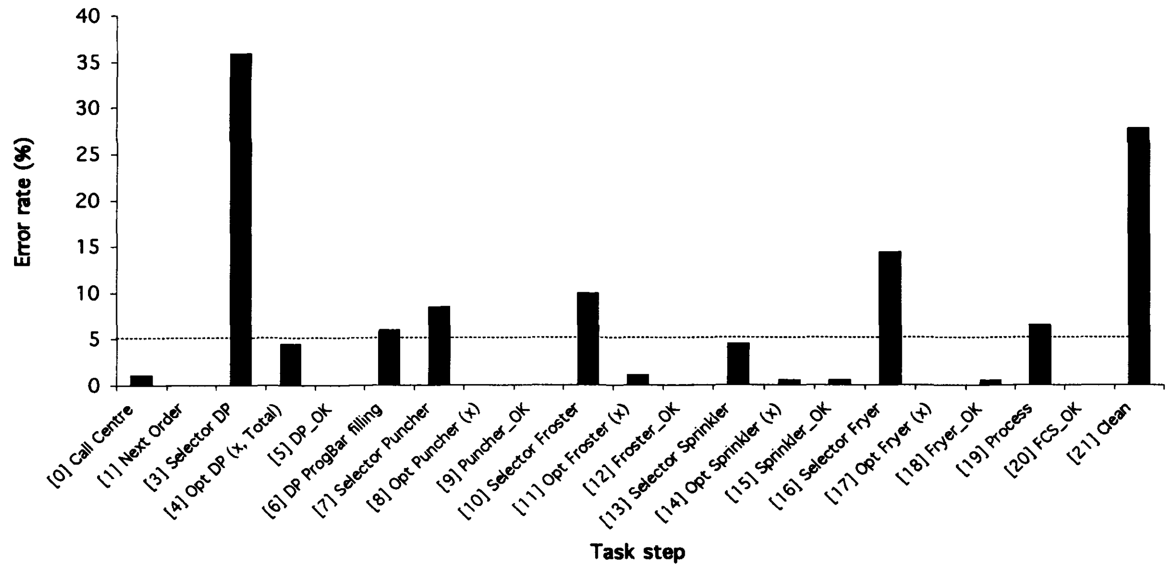


Figure 6.13: Error rate of each task step in PC trials.

The pattern of errors occurring across the task steps is consistent with the pattern from Experiment 6b; errors at the "Selector DP" have the highest error rate (about 36%), followed by errors at the PC step ("Clean"; about 27%). Other errors that have an error rate exceeding the 5% criterion are mostly occurring at one of the "Selector" steps. Error at step [6] also reached systematicity but, just as in the previous experiments, this error is not to be further analysed because it is symptomatic of participants being impatient: not waiting long enough for the progress bar to fill up before carrying out the next action. The nature of this error is not of particular interest to the current thesis.

3.1.2 Interruption position effect on PCE

Table 6.19 shows the number and error rate of PCE in relation to the different interruption positions. Each participant received 2 trials for each interruption position; the total number of opportunities are computed across all 34 participants.

	Interruption position		
	Z	Other	Nil
Total no. of PCE (Total no. of opportunities)	28 (68)	12 (68)	16 (68)
Mean error rate (SD)	41.2% (43.5)	17.6% (30.0)	23.5% (30.7)

SD = Standard deviation

Table 6.19: Number of PCE and their mean error rates with respect to the different interruption positions.

A one-way repeated ANOVA yielded a large significant main effect of interruption position, $F(2, 66) = 4.938$, $p = .01$, Eta squared = .13. Planned comparisons between positions Z and Other showed a reliable difference, $t(33) = 2.61$, $p = .014$, and difference between positions Z and Nil was also significant, $t(33) = 2.244$, $p = .032$. Comparison between positions Other and Nil did not show a reliable difference, $t(33) = .941$, $p = .353$. The overall pattern of results is consistent with the previous experiments.

3.1.3 PCE and non-PCE resumption pattern

There was one error occurring at the PC step after interruption Z, which involved omitting the PC step but the resumed action was not at the Call Centre, instead it was at one of the steps in the doughnut-making task. This error was not classified as a PCE, but as a non-PCE for later analysis, because it did not follow the operational criteria which required responding to the Call Centre when the PC step was omitted. All other PCEs that occurred immediately after interruption Z had the same pattern of resumption: they were errors omitting the PC step and moving on to the Call Centre task.

In contrast there was a qualitatively different resumption pattern in the non-PCEs that occurred immediately after an interruption; 38% (13 out of 34) involved resuming to the task step just before the interruption, and the remainder lacked a consistent pattern of resumption. The trends of the results are consistent with those obtained in the previous experiment. This further confirms that the cues in the task environment play a part in resuming an interrupted task.

3.1.4 Resumption times between different task steps

Just as in previous experiments, the analysis of resumption times was carried out on correct resumption data only to avoid potential biases from incorrect resumption behaviour. Valid data from 15 participants were subjected to analysis, these participants resumed correctly at least once for the PC step and the non-PC steps.

Table 6.20 shows the mean resumption times for PC and non-PC steps. The same pattern of results was obtained as the previous experiments; resumption times between the task steps was found to be significantly different, $t(14) = 4.365$, $p = .001$.

	PC step	Non-PC step
Mean resumption time	3.4s	7.1s
(SD)	(1.2)	(2.7)

Table 6.20: Mean resumption times after interruption Z and Other.

Like the analysis in the previous experiment, movement times estimated using Fitts' Law were taken out from the resumption times to avoid possible confounds. The same movement times from the previous experiment were used because the screen and button layout are identical in the current experiment. Table 6.21 below shows the mean resumption times with the movement times taken out.

	PC step	Non-PC step
Mean resumption time	3.2s	6.4s
(SD)	(1.2)	(2.7)

Table 6.21: Mean resumption times (minus estimated movement times) after interruption Z and Other.

The difference between the task steps was still found to be significant, $t(14) = 3.92$, $p = .002$.

3.1.5 Interruption position effect on non-PCE

There were a total of 189 non-PCEs and they were categorised into "immediately after interruption", "later after interruption" and "no interruption" (including errors before an interruption). The procedures for counting the error occurrences and calculating their respective error opportunities are the same as the previous experiment. Table 6.22 shows the number and mean error rates in relation to the different interruption position. About half of the errors in the "no interruption" category (73 out of 144) consist of errors occurring at "Selector DP" (step [3]). For the same reason as the previous experiment's analysis, these errors and their error opportunities are excluded from the following analysis.

One of the non-PCEs was a resumption error at the PC step after interruption Z, as mentioned earlier, this error is excluded from the analysis to avoid confounding the calculation of error opportunities in the "immediately after interruption" category. The occurrence of this error will be discussed later in the chapter.

	Interruption position			Total no. of non-PCE	
	Immediately after	Later after	No interruption		No interruption*
Total no. of non-PCE (Total no. of opportunities)	34 (68)	10 (372)	144 (3436)	188	71 (3232)
Mean error rate (SD)	50.0% (44.4)	3.1% (6.8)	4.2% (3.5)		2.2% (2.4)

**This "no interruption" category excludes errors at "Selector DP" step*

Table 6.22: Number of non-PCEs and their mean error rates with respect to the different interruption positions.

A one-way repeated ANOVA was performed and obtained a large significant main effect. The data did not conform to the assumption of sphericity; therefore, the Greenhouse-Geisser correct was used; $F(1.028, 33.918) = 40.875$, $p < .001$, Eta squared = .553. Post hoc comparisons with Bonferroni correction showed a reliable difference between "immediately after" and "later after", and "immediately after" versus "no interruption" also yielded a significant difference. There was no reliable difference between "later after" and "no interruption". The result of interruption position effect on non-PCEs is consistent with the finding from the previous study.

3.1.6 Categorisation of non-PCE

The same analysis on the obtained non-PCEs was carried out as in the previous experiments. The identification of the errors was based on the first mouse-click at a task step. The same four categories of errors (skip-selector, incorrect sequence, skip-and-incorrect and miscellaneous) were identified as in the previous study and their number of occurrences and relative proportions are summarised in Table 6.23. Consistent with the pattern of results of the previous experiment, the majority of the non-PCEs were selector-step errors. The relatively small proportion of miscellaneous errors consists of errors occurring at other task steps.

	Selector-step errors			Miscellaneous errors
	Skip-selector error	Incorrect-sequence error	Skip-and-incorrect error	
No. of errors	113	44	2	30
Proportion	60%	23%	1%	16%

Table 6.23: The relative proportion (%) of the different identified error kinds.

3.1.7 Interruption position effect on selector-step errors

Like the analysis carried out in the previous study, the obtained selector-step errors are subjected to finer detail analysis according to the identified error categories. The logic and procedures of error count and error opportunities calculations are the same as the previous experiment.

3.1.7.1 Skip-selector errors

More than half of the skip-selector errors (68 out of 113) occurred at "Selector DP" and these errors are excluded from the following analysis, for the same reason as the previous studies, because action executed on this step was not preceded by the interruption manipulation. Table 6.24 shows the number and mean error rate of skip-selector errors with respect to the interruption positions.

	Interruption position		
	Immediately after	Later after	No interruption
Total no. of Skip-selector errors (Total no. of opportunities)	1 (49)	7 (73)	37 (692)
Mean error rate (SD)	1.5% (8.6)	11.5% (24.1)	5.2% (7.6)

Table 6.24: Number and mean error rates of skip-selector errors according to different interruption position.

A one-way repeated ANOVA was performed and found no significant main effect of interruption position, $F(1.297, 42.791) = 3.093$, $p = .076$ (with Greenhouse-Geisser correction).

3.1.7.2 Incorrect-sequence errors and PCEs

Like the analysis of the previous study, skip-and-incorrect errors were also included in the analysis of incorrect-sequence errors; yielding a total of 45. There were 5 incorrect-sequence errors occurring at "Selector DP" step, and they are excluded from the analysis. Furthermore, as mentioned earlier, there was one resumption error occurring at the PC step and it is also excluded from the analysis. Table 6.25 shows the number and mean error rates of incorrect-sequence error and PCE in relation to the interruption positions.

	Interruption position		
	Immediately after	Later after	No interruption
Total no. of Incorrect-sequence errors (Total no. of opportunities)	33 (68)	0 (124)	7 (828)
Mean error rate (SD)	48.5% (45.2)	0% (0)	0.9% (2.2)
Total no. of PCE* (Total no. of opportunities)*	28 (68)	12 (68)	16 (68)
Mean error rate* (SD)*	41.2% (43.5)	17.6% (30.0)	23.5% (30.7)

* *These figures are the same figures from Table 6.19*

Table 6.25: The number and mean error rates of incorrect-sequence errors and PCE according to the different interruption positions.

A 2 x 3 (error type x interruption position) repeated ANOVA was performed on the error rates of incorrect-sequence error and PCE. A large reliable main effect of interruption position was found, $F(1.439, 47.501) = 30.1$, $p < .001$, Eta squared = .477 (with Greenhouse-Geisser correction). A moderate significant main effect of error type was also found, $F(1, 33) = 5.588$, $p = .024$, Eta squared = .145. The interaction between error type and interruption position was also found to be significant with a moderate effect size, $F(1.421, 46.905) = 5.381$, $p = .015$, Eta squared = .14 (with Greenhouse-Geisser correction).

A nonparametric test was used in analysing the effect of interruption position on incorrect-sequence error because of the occurrence of zero variance in one of the interruption position cells. Post hoc comparisons using Wilocoxon Signed Rank test, with Bonferroni corrections, obtained a reliable difference between "immediately after" versus "no interruption".

The results suggest that interruption position has the same general effect on incorrect-sequence error as on PCE, which is consistent with the findings from the previous experiments. The significant interaction between incorrect-sequence error and PCE is likely to be due to the error rate difference between the two error types in the "later after" and "no interruption" categories. In the case of PCE, the mean error rates in both interruption categories ("later after" and "no interruption") are higher than those in incorrect-sequence error. This suggests PCE has a generally higher rate of occurrence than incorrect-sequence error when it is not preceded immediately by an interruption.

3.1.8 Errors at "Selector DP" step

A total of 73 errors occurred at the very first compartment step — "Selector DP". The same consistent pattern as the previous experiment was obtained: that over 90% (68/73) of the errors at "Selector DP" step were skip-selector errors and the remaining were incorrect-sequence errors.

3.2 Analysis of non-PC trials

3.2.1 Overall errors

Among the non-PC trials, there were 368 errors in total across the 34 participants. There were 12 non-PC trials for each participant; therefore, each task step has 408 opportunities for an error to occur. Figure 6.14 below shows the distribution of the error rates across all task steps[9].

[9] The numbering of the task steps is slightly different from the one in the PC trials because there was not an extra step (acknowledging the PC step modal dialog box) after "Next Order" (step [1]) in the non-PC trials.

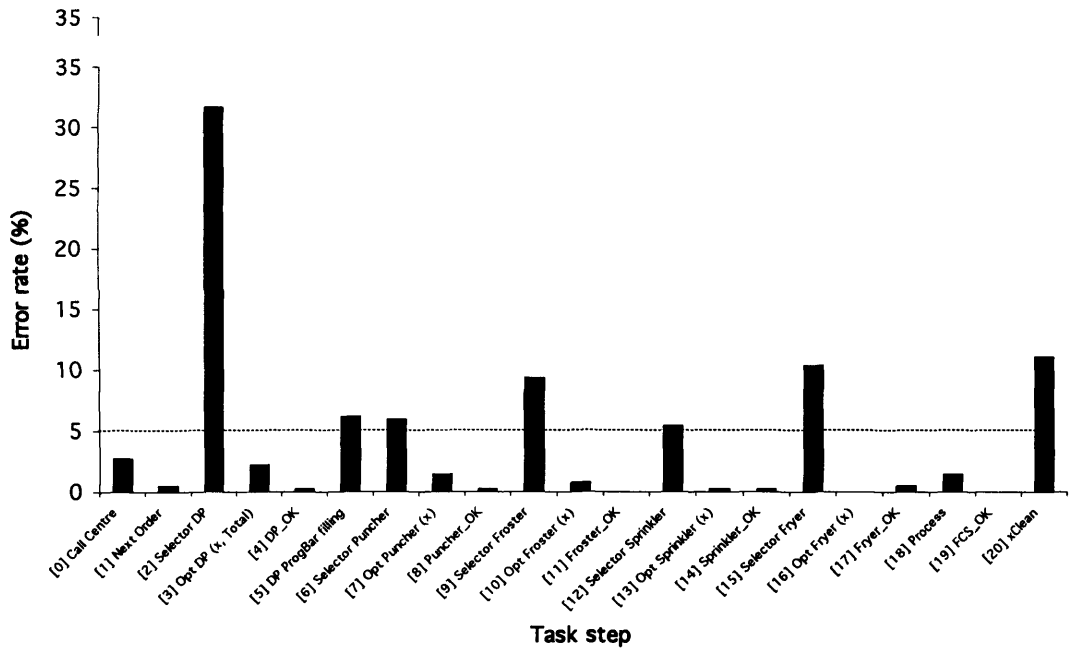


Figure 6.14: Error rate of each task step in non-PC trials.

The pattern of error rates across task steps is the same as the PC trials; errors that reached the 5% systematicity criterion were almost all occurring at the "Selector" steps with "Selector DP" having the highest error rate (about 31%). Error at step [5] also has an error rate above the criterion but its nature is not overly interesting as mentioned earlier.

Error at the PC step ("xClean") also appears systematic; although they occur at the PC step, they are not PCE as such according to the definition operationalised in the task. These errors can be thought of as "extra Cleaning" errors as they were the execution of the "Clean" step that was not required in the first place. There were 45 xClean errors in total and table 6.26 shows the number and mean error rates in relation to the interruption positions.

	Interruption position		
	Z	Other	Nil
Total no. of xClean (Total no. of opportunities)	8 (68)	11 (68)	26 (272)
Mean error rate (SD)	11.8% (24.8)	16.2% (26.7)	9.6% (16.9)

Table 6.26: Number of xClean and their mean error rates with respect to the different interruption positions.

A one-way repeated ANOVA on the error rate did not find a significant effect of interruption position, $F(2, 66) = 1.129$, $p = .329$. The occurrence of these errors is discussed in more detail later in the chapter.

3.2.2 Interruption position effect on non-PCE

There were a total of 323 non-PCEs and they were categorised into "immediately after interruption", "later after interruption" and "no interruption" (including errors before an interruption). Table 6.27 shows the number and mean error rates for the non-PCEs according to the different interruption positions. 7 of the non-PCEs were resumption errors at the PC step after interruption Z. The error is excluded from the analysis to avoid confounding the calculation of error opportunities in the "immediately after interruption" category. The occurrence of this error is discussed later in the chapter.

	Interruption position			Total no. of non-PCE	
	Immediately after	Later after	No interruption		No interruption*
Total no. of non-PCE (Total no. of opportunities)	38 (68)	8 (414)	270 (7261)	316	143 (6853)
Mean error rate (SD)	55.9% (36.4)	1.6% (3.3)	3.7% (2.5)		2.1% (1.4)

**This "no interruption" category excludes errors at "Selector DP" step*

Table 6.27: Number of non-PCEs and their mean error rates with respect to the different interruption positions.

Almost half of the errors in the "no interruption" category (127 out of 270) are errors occurring at "Selector DP" (step [3]); for the same reason as the previous analysis on PC trials, these errors and their error opportunities are excluded from the analysis. A one-way repeated ANOVA revealed a large significant interruption position effect, $F(1.012, 33.402) = 74.041$, $p < .001$, Eta squared = .692. Post hoc comparisons with Bonferroni correction yielded a reliable difference between "immediately after" and "later after"; the difference between "immediately after and "no interruption" was also found to be reliable. No significant difference was found between "later after" and "no interruption".

3.2.3 Categorisation of non-PCE

The non-PCEs were categorised into the same categories identified in the previous analysis on the PC trials. Table 6.28 shows the number and the relative proportion of each error category. The same pattern of error proportions as the PC trials was obtained; most errors were selector-step errors and a relatively small proportion of errors occurred at other steps in the task.

	Selector-step errors			Miscellaneous errors
	Skip-selector error	Incorrect-sequence error	Skip-and-incorrect error	
No. of errors	202	51	5	65
Proportion	63%	16%	1%	20%

Table 6.28: The relative proportion (%) of the different identified error kinds.

3.2.4 Interruption position effect on selector-step errors

Like the previous analysis, selector-step errors were analysed in relation to interruption position. Categorisation of errors into the different interruption positions adopted the same logic and procedures as before.

3.2.4.1 Skip-selector errors

More than half of the skip-selector errors (124 out of 202) occurred at "Selector DP" and these errors are excluded from the following analysis for the same reason as the analysis on PC trials. Table 6.29 below shows the number and mean error rate of skip-selector errors with respect to the interruption positions.

	Interruption position		
	Immediately after	Later after	No interruption
Total no. of Skip-selector errors (Total no. of opportunities)	0 (58)	6 (83)	72 (1491)
Mean error rate (SD)	0% (0)	6.3% (16.1)	4.8% (5.9)

Table 6.29: Number and mean error rates of skip-selector errors according to different interruption position.

A nonparametric test was used in analysing the effect of interruption position on skip-selector error because of the occurrence of zero variance in one of the interruption position cells. Post hoc comparisons using Wilcoxon Signed Rank test, with Bonferroni correction, revealed no reliable difference between "later after" and "no interruption". As there was no error occurrence in "immediately after"; therefore, this is still consistent with the previous finding on skip-selector error: its occurrence is not affected by an immediately preceding interruption.

3.2.4.2 Incorrect-sequence errors

Skip-and-incorrect errors were also included in the analysis of incorrect-sequence errors; yielding a total of 56 errors. Three incorrect-sequence errors occurring at the "Selector DP" step were excluded from the analysis. Furthermore, there were 7 resumption errors occurring at the PC step and, as in the previous analysis, they were also excluded from the analysis. Table 6.30 shows the number and mean error rates of incorrect-sequence error and PCE in relation to the interruption positions.

	Interruption position		
	Immediately after	Later after	No interruption
Total no. of Incorrect-sequence errors (Total no. of opportunities)	38 (68)	0 (141)	8 (1831)
Mean error rate (SD)	55.9% (36.4)	0% (0)	0.4% (0.8)

Table 6.30: The number and mean error rates of incorrect-sequence errors and PCE according to the different interruption positions.

A nonparametric test was used in analysing the effect of interruption position on incorrect-sequence error because of the occurrence of zero variance in one of the interruption position cells. Post hoc comparisons using Wilcoxon Signed Rank test, with Bonferroni correction, revealed a reliable difference between "immediately after" versus "no interruption". The result is consistent with the previous findings that there is more likely to be an incorrect-sequence error immediately after an interruption than after an earlier interruption or no interruption at all.

3.2.5 Errors at "Selector DP" step

A total of 127 errors occurred at the very first compartment step — "Selector DP". The same consistent pattern as the previous experiment was obtained; of all the errors at "Selector DP", over 95% (124/127) were skip-selector errors and the remaining were incorrect-sequence errors.

4 Discussion

The current experiment has also successfully generated an overall level of PCEs (about 27%; 56 out of 204 opportunities) that allows further investigations into the error occurrence with respect to the effect of interruption. This further confirms that the design of the current task is a successful paradigm in provoking PCEs in an experimental setting. Furthermore, a relatively higher overall PCE rate than Experiment 6b (about 16%; 31 out of 341 opportunities) is generated with the inclusion of a conditional PC step.

The main finding of the effect of interruption position on PCE has been replicated in the current experiment: that an interruption is more likely to result in a PCE when it occurs *just before* the PC step than earlier in the task or with no interruption at all. Moreover, the difference in resumption times between the PC step and non-PC steps was also replicated. Resuming the PC step after an immediate interruption (i.e. Z) was faster than resuming earlier non-PC steps immediately after an interruption (i.e. Other). This further confirms that cues in the task environment facilitate resumption of an interrupted task.

Further analysis of other errors for both PC trials and non-PC trials replicated the same effect of interruption position; that an error is more likely to occur immediately after an interruption than later after or with no interruption at all. Finer grain analyses on errors occurring on the selector steps suggest consistent results across both PC and non-PC trials: skip-selector errors are not sensitive to the disruptive effect of an immediately preceding interruption, whereas incorrect-sequence errors are more likely to occur immediately after an interruption. These same patterns of results for skip-selector errors and incorrect-sequence errors were also obtained in Experiments 6b and 6c.

The interaction between error type (PCE and incorrect-sequence error) and interruption position was found to be significant, which is in contrast to the results of Experiment 6b. While the interaction effect might suggest the effect of interruption position has a differential effect on PCE and incorrect-sequence error, interpretation of the result needs caution. The statistically significant interaction effect is likely to be due to the general increase in PCE rate in the current experiment. The average PCE rates for interruption positions Other and Nil is about 20% in the current experiment, whereas, in Experiment 6b the average PCE rate for the two interruption positions is about 8.7%. The increased PCE rate in interruption positions Other and Nil in the current experiment might serve as an elevated baseline for comparing the error rate immediately after an interruption (position Z). As a consequence, the disruptive effect immediately after an interruption might be reflected as less pronounced than in Experiment 6b where there is a lower baseline PCE rate. However, what is clear in the results of the current experiment and Experiment 6b is that both error types — PCE and incorrect-sequence error — are sensitive to the disruptive effect of an immediately preceding interruption.

In the current design of the experiment, another type of error — xClean — was obtained at the PC step in the non-PC trials. The occurrence of this error is a result of executing the PC step when not required to do so. The error appears to be systematic in terms of its occurrence. Furthermore, the error was not found to be sensitive to the disruptive effect of an immediately preceding interruption. The occurrence of the error suggests misremembering, or confusion, of an occasionally required action.

The original objective of the task modification in the current experiment was to imitate the less-practised PC step characteristic of Byrne & Bovair's cyclic task structure in order to obtain a higher overall PCE rate. However, the current modification is slightly different to Byrne & Bovair's cyclic task structure. Instead of cycling through a task procedure a number of times before completion as in Byrne and Bovair's task, participants in the current experiment were asked at the beginning of some trials to carry out the PC step. In retrospect, the modified task has a different element to it, namely a prospective memory element, concerned with whether one remembers to carry out some actions in the future when the appropriate time comes. Arguably, a conditional-PC step seems to relate to prospective memory in that both require remembering to perform some actions in the future. The nature of the conditional-PC step is discussed in more detail in the next section.

In relation to the current experiment, the occurrence of the xClean error is likely to be the result of extra memory load imposed by the prospective nature of the task causing misremembering of the action. Similarly, the relatively higher overall PCE rate obtained in the PC trials compared to Experiment 6b is also likely to be due to the extra memory demand imposed by the conditional PC step. While the PCEs obtained in the current experiment might include occasions when participants misremembered the PC action (whether to carry it out or not) rather than genuinely forgetting to carry it out, it is unlikely that all PCEs obtained are due to misremembering or confusion about the required action. This is because the omission of the PC step in the current experiment is still found to be sensitive to the effect of interruption position, as was found for PCEs obtained in the previous experiments. However, the xClean error, which is suggested to be due to misremembering, is not found to be sensitive to the effect of interruption position.

The resumption pattern at the PC step involved a few errors carrying out other actions rather than moving on to the Call Centre, whereas such errors did not occur in the previous experiment. In the PC trials, there was one such error (less than 1 %) and in the non-PC trials there were 7 (about 1%). Although the occurrence of the error did not appear systematic, as a speculation of the presence of the error, the occasional requirement of the PC action might add more memory demands to the PC step resulting in other unpredicted erroneous actions.

Consistent results were obtained between PC and non-PC trials in terms of errors at the "Selector DP" step: almost all of the errors at that step were skip-selector errors and only a few were incorrect-sequence errors. This further confirms the previous finding of dissociation between correct task sequence execution and correct usage of the interface.

Overall chapter discussion

By adopting a similar paradigm to Byrne & Bovair (1997) using a procedural task with a fixed task sequence, the current series of experiments has successfully generated PCE rates that are high enough for investigations in a laboratory setting. The low PCE rate obtained in Experiment 6a led to a comparative analysis between the Doughnut task and the Phaser task used in Byrne & Bovair's study. The analysis suggested several differences between the two tasks that might have contributed to the low error rate generated from the Doughnut task. In Experiment 6b, the Doughnut task was modified to include a competing signal upon task completion and a smaller task (Call Centre) was implemented so that one had to physically turn away from the main Doughnut task in order to attend to it. Furthermore, the button for the PC step was also modified to appear less prominent than originally designed. The inclusion of these modifications in Experiment 6b led to a generation of 16% of the overall PCE rate, which was high enough to examine the effect of interruption position statistically. In Experiment 6d, the task was further modified to make the PC step a less practised step relative to other steps in the task. The purpose of Experiment 6d was not only to replicate findings from the previous experiment but also to mimic one of the characteristics in Byrne and Bovair's Phaser task.

Given the scant research on PCE, the task paradigm designed and used in this study has a significant contribution to the task repertoire in studying the error phenomenon. More importantly, the modifications identified and made in the current series of experiments reveal not only some important factors in relation to inducing the error in the laboratory, but also some properties inherent to PCE. One such property that differentiates PCEs from other omission errors is the occurrence of a false completion signal upon task completion, which could act as a competing signal against the PC step. Everyday PCE examples in procedural tasks, such as photocopying and withdrawing cash from the ATM, all involve occurrence of a competing signal when the task goal is fulfilled: having the copies or the cash in your hand. This task completion signal is "competing" with the PC step because it has a tendency to urge one to move on to the next task goal without remembering to perform the secondary PC step.

The series of experiments investigated the effect of interruption, more specifically the effect of interruption position and duration, on the rate of PCE in a procedural task. Consistent results were produced in Experiments 6b, 6c and 6d, showing that the rate of PCE could be affected by manipulating *where* the interruption occurs in the primary task. The findings are consistent with the predictions made from Altmann and Trafton's AGM model, that interruptions occurring just-before the PC step in a task were more likely to result in a PCE than earlier interruptions due to the disruption of associative priming from its previous task step. During the interrupting activity, rehearsal of the to-be-resumed goal is prevented and its level of activation undergoes a time-based decay process. When the interrupting activity is long enough for the to-be-resumed goal to decay, an incorrect goal is likely to be retrieved upon task resumption.

The time it took the participants to resume the primary task after an interruption was also found to be different depending on where the interruption occurred. A consistent pattern of results was found in all three experiments that resumption times were longer when one was to resume any of the major task steps before the task was completed, whereas resumption times were shorter for the PC step after an interruption. This is likely to be due to the relatively lower encoding demand of the PC step compared to the other major task steps resulting in faster resumption. The results are consistent with previous findings (Monk et al., 2002; 2004) suggesting that the encoding demand before completion of a subtask is higher than at the beginning of a subtask or in a repetitive operation. The higher encoding demand results in longer resumption time because more information needs to be rehearsed about the goal state and its associative links to the next task step.

The effect of interruption position on PCE has a commonality to Chung & Byrne's (2004) finding, though they examined the role of dynamic visual cues in combating PCEs. It was found that an effective visual cue to combat PCE had to come into play just in time, i.e. just-before the PC step, and that cues occurring early on in the task did not aid the reduction of the error. Although the current study and Chung & Byrne investigated different manipulations on the error phenomenon, both studies grounded their theoretical predictions in the AGM model highlighting the importance of timing of the manipulations in relation to PCE.

Although PCE was found to be sensitive to the effect of interruption position, analyses of other errors in Experiments 6b, 6c and 6d suggest that the effect is not unique to PCE. Finer grained analyses of other procedural errors suggest that it is more likely to result in an error of resuming the correct task step in a sequence (incorrect-sequence error) after an immediately preceding interruption than an earlier interruption. The analyses also reveal the presence of another kind of omission error: incorrect interface usage (skip-selector errors) but correct task sequence execution. These skip-selector errors, however, are found not to be sensitive to the disruptive effect of an immediately preceding interruption. The distinction between incorrect-sequence error and skip-selector error suggests very different manifestation in terms of error behaviour. When an incorrect-sequence error is committed, the participant is executing a wrong sequential step in the procedure but maintaining the usage of the interface correctly (not forgetting to "select" one of the chosen compartments in the Doughnut Machine). In contrast, the commission of a skip-selector error suggests that the participant is remembering the sequence of the task procedure correctly but using the interface incorrectly. Moreover, these two kinds of error respond to the disruptive effect of an immediately preceding interruption very differently: an incorrect-sequence error is more likely to occur immediately after an interruption but this is not the case for skip-selector error. The commonality between incorrect-sequence error and PCE is that their occurrences are increased when preceded by an interruption. This might imply that knowledge of a correct task sequence or a PC step during task execution has a dynamic nature, which is susceptible to disruption caused by an immediately preceding interruption. Whereas knowledge of correct interface usage might have a relatively stable nature during task execution and so is less vulnerable to the disruption caused by an immediately preceding interruption. This differentially disruptive effect of interruption on different error kinds provides support to the general notion that occurrence of PCE is not due to deficiency in one's knowledge of how to use the interface but momentary memory lapses caused by factors such as interruption or high working memory demand (Byrne & Bovair, 1997).

Analyses of other errors also allow comparison between PCE and incorrect-sequence error in terms of the effect of interruption position. However, it is difficult to conclude from the current series of experiments whether one kind of error is more disrupted by an interruption than the other. This is because there is no interaction effect (error type × interruption position) obtained in Experiment 6b but there is in Experiment 6d. Although the significant interaction effect yielded in Experiment 6d is attributed to an overall increase in PCE rate, consistent across the current experiments is the finding that both PCE and incorrect-sequence error are sensitive to an immediately preceding interruption. However, the question regarding whether the extent of disruption differentiates between the two kinds of error would require further research to explore it fully.

Results from Experiment 6c show that PCEs are sensitive to the disruption effect of interruption as brief as 15 seconds. The manipulation used in the experiment does not show the predicted interaction effect between interruption position and duration. However, the trend of the data is consistent with the predicted direction: overall fewer PCEs occur in the 15-second condition than the 45-second condition.

Although the interruption duration manipulation does not yield the predicted effect on PCEs, the shorter interruption is found to have a less disruptive effect on another kind of error, namely, incorrect-sequence error; showing an interaction effect between interruption position and duration. The effect of interruption position on incorrect-sequence error persists in the 45-second condition but not in the 15-second condition; in other words, the error is increased when immediately preceded by a longer interruption but is less affected by a shorter interruption. As mentioned earlier, while consistent results from the current series of experiments suggest that both PCE and incorrect-sequence error are increased by an immediately preceding interruption, manipulation of interruption duration reveals different behaviour between the two kinds of error. Current results suggest that PCE is sensitive to interruption as short as 15 seconds but incorrect-sequence error is not. Early research on interruption duration (e.g. Gillie & Broadbent, 1989) suggests no differential effect on primary task performance; however, results from the current study suggests the contrary: different durations can have differential effect on different kinds of error. Using more local dependent measures, such as number of errors or resumptions times (Trafton et al., 2004), differential disruptive effect imposed by different interruption durations can be revealed.

The overall PCE rate obtained in Experiment 6d is about 10% more than that obtained in Experiment 6b and this is attributed to the inclusion of a conditional PC step. The characteristic of a conditional PC step is that it only requires execution on some occasions within a task, and this is different to the PCEs observed in most everyday situations, such as retrieving your original document after photocopying, which requires the execution of the PC step every time the main goal is completed. On reflection, although the PC step in both Experiment 6d and Byrne and Boviar's Phaser task is less frequently practised relative to other steps in the tasks, the nature of the conditional-PC step manipulation is different to the characteristic of a cyclic task structure in the Phaser task. This is because the appearance of the condition to carry out the PC step occurs at the very beginning of the task in Experiment 6d (notification pops up before the participant could begin the doughnut-making process), whereas the PC condition occurs at the very end of the Phaser task in Byrne and Bovair's experiment: a participant is given feedback about whether an enemy ship is destroyed *after* various configuration processes of the Phaser machine are done.

In retrospect, the conditional-PC step manipulation in the current study contains an element of prospective memory because the participant is notified about the requirement of the PC step at the beginning of some trials, and he/she has to carry out that step when the appropriate time arrives: in this case, at the end of the task procedure. The prospective memory nature of the manipulation is suggested to have added memory load to performing the PC step, because the participant needs to remember whether he/she has received the notification in a given trial. It is possible that the omissions of the PC step in Experiment 6d might not necessarily reflect the occurrences of PCEs entirely. It is speculated that the failure to execute the conditional-PC step might be due to two reasons: first, the participants genuinely forgot to carry out the PC step, thinking one has reached the end of the task and this can be classified as a PCE. Secondly, the participants might have actually remembered about the execution of the PC step but did not carry out the required action because they misremembered whether the PC step had to be carried out or not. In this situation, the omission of the PC step is not necessarily a PCE. This suggests, in Experiment 6d, the overall increased error rate at the PC step might contain the error due to misremembering as well as occurrences of PCE.

The extra memory imposed by the conditional-PC step is also reflected in the occurrences of a commission error — xClean (execution of the PC step when it is not required) — due to misremembering of information. Furthermore, the analysis of interruption position on the error suggests that it is not sensitive to an immediately preceding interruption. However, the occurrence of the commission error is not an intended outcome of the current study and the exact nature of the error would require further investigation.

Traditional research on prospective memory and PCE appear to be rather separate and this is primarily because of two reasons: first, experimental tasks used to study prospective memory are not usually procedural tasks defined by a set of fixed procedures as is the case for PCE. Second, a theoretical distinction made is that a prospective memory error is concerned with failure to carry a formed intention and prospective memory researchers view PCE not being under such distinction because the PC step constitutes only a small part of an overall intention (Kvavilashvili & Ellis, 1996). Speculations made on the conditional-PC step manipulation in Experiment 6d suggest that the distinction between prospective memory and PCE might be less clear than one used to think. However, this suggestion remains speculative and further research is required in order to understand the relation between prospective memory and PCE, or better still, if it is possible to make novel experimental or theoretical predictions about PCE from the perspective of prospective memory research.

There was another highly systematic error found in the experiments, namely, skip-selector error at the very first major step of the task ("Selector DP"). Initial results from the current study suggest that the skip-selector error is not sensitive to the disruptive effect of interruption. Another intriguing aspect of the error is its highly systematic occurrences at the very first step of the main task process. The systematic omission at the beginning of a sub-task (forgetting to 'select' a compartment before operating on it) suggests its role in the task sequence being secondary rather than primary. In other words, the subgoal of having to 'select' a to-be-operated compartment does not occur as a natural step in the task sequence. The structure of the error bears similarity to a PCE in that it involves omitting a "secondary" step in a procedure; on the other hand, it is different to a PCE in that it occurs at the beginning of a task rather than at the end when the task is completed. At the same time, a skip-selector error is extremely similar to, if not the same as, the USW error which occurs when one forgets to select a window before interacting with it (Lee, 1992). Further research is necessary in order to investigate the nature and systematic occurrences of the error.

There are a number of methodological issues stemming from the series of experiments. First of all, in Experiment 6a, the between-subject interruption conditions (complex vs simple) utilised two interruptions of different nature and durations. The limitation of this is the variation of two attributes rather than a single one. However, the objective of Experiment 6a is to set up a methodological paradigm to ensure the 75-sec mental arithmetic interruption has a disruptive effect. The use of a simple perceptual-motor task as a short interruption is to serve as a comparative basis for the longer and more complex interruption on its disruptiveness. The use of a seemingly different interruption was not to investigate and draw conclusions about the effects of the nature and duration of interruptions as such.

Secondly, one might criticise the categorisation of the non-PCEs into the interruption position categories as not comparing like for like when compared to the PCEs in Experiment 6b and 6d. More specifically, the "no interruption" category containing errors from Nil trials as well as from errors before an interruption for non-PCEs. However, this particular grouping is argued to be equivalent, at least functionally, to the Nil category for PCEs in order to analyse the obtained data in an exhaustive manner.

All in all, the effect of interruption on PCE is clear and robust: an immediate preceding interruption is more likely to result in a PCE than an earlier interruption. However, further research is required to confirm the effect of the interruption duration on PCE, by including a shorter interruption duration than the current 15 seconds to examine the predicted interaction effect between interruption position and duration. The AGM model serves as a good theoretical framework allowing predictions to be made about the effect of interruption. Traditional studies in interruption research did not look at how an interruption might affect specific kinds of error. The current series of experiments has furthered the understanding of the occurrences of PCE in a procedural task by investigating the effect of interruption with an appropriate theoretical framework.

The next section takes the current general findings of the disruptive effect of interruption on PCE and discusses how the various existing theoretical approaches to the error might or might not apply to the findings.

Theoretical discussion

The main findings from the series of experiments investigating the effect of interruption position on PCE show that the error is more likely to occur immediately after an interruption than later after or with no interruption at all. The effect, however, was found not to be unique to PCE but to affect other non-PCEs as well, more specifically incorrect-sequence errors (errors executing the incorrect task step). The incorrect resumption pattern of PCE also highlights that the incorrectly resumed step is, mostly, the next task goal in sequence and not just any random task step.

As discussed earlier in Chapter 2, the major shortcoming in the PCE account offered by the AGM framework is that it better explains how PCE is avoided rather than how the error occurs. Nevertheless, the framework provides low enough level of specificity to generate predictions about the effect of interruptions on the error. The AGM model explains the interruption position effect on PCE in terms of three basic notions: first, the mechanism of associative cueing in a procedural task; secondly, reliance of cues in the environment upon resuming the interrupted task; thirdly, the time-based decay of goal memory. An interruption disrupts the associative links in memory between the PC step and its preceding step, and during the interrupting activity the activation of the next correct subgoal undergoes a decay process. Upon task resumption one relies on cues in the environment to cue the next task step. In situations where interruption occurs just before the PC step, a false completion signal is the main cue prompting one to move on to the next task upon task resumption. As a result, the PC step is omitted. To put it simply, associative links in memory cue the PC step, whereas environmental cues (the false completion signal) prime the next task goal in sequence. The same explanation can also be applied to other errors (e.g. incorrect-sequence errors) during task execution in relation to the effect of interruptions. However, the AGM model is not able to explain the unanticipated finding that PCEs are more vulnerable to the effect of interruption than incorrect-sequence errors.

Chung & Byrne (2004) also adopted the AGM framework to generate predictions about the use of dynamic reminders in combating PCE. Although Chung & Byrne's study had the opposite intention to our series of interruption experiments, which investigate factors provoking the error, empirical findings from both studies highlight the just-in-time factor of the variables manipulated. An effective dynamic reminder was found to be effective only if it occurred just-before the PC step. An interruption occurring just-before the PC step was found to be more detrimental to the error than interruptions occurring at other steps in a task. The AGM framework offers adequate theoretical guidance in explaining the two sets of empirical findings.

Byrne & Bovair's CAPS account of PCE has a rather different decay process to the AGM accounts. Instead of a time-based decay process, the process is load-based in CAPS which accounts for PCE as a function of working memory demand. It is suggested that the occurrence of PCE is "...not simply as a result of delay between when the main goal is satisfied and when the post-completion step should be executed." (p. 41). This fundamental difference seems problematic in explaining the effect of interruption on PCE, which depends on the decay of the PC subgoal during the length of the interrupting activity. Despite the difference there are other aspects in the CAPS model suggesting that it may be stretched to accommodate the effect of interruption on the error.

In CAPS, activation is spread and maintained from the main goal to its subgoals as long as it remains active in memory. When the main goal is completed it gets displaced from memory and its subgoals stop receiving activations from it. However, under low working memory load, the main goal will remain active in memory for some time upon completion and productions for maintaining active goals will fire. This is to propagate activation to the PC subgoal so that it gets executed. According to this view, the occurrence of an interruption can be seen as replacing the main goal of the primary task in memory with the interrupting activity. When an interruption occurs just before the PC step, the main goal of the primary task had already been completed. The interrupting activity prevents the firing of the goal maintenance productions. As a result, upon task resumption, the PC subgoal would not have received enough activation to be executed. On the other hand, interruptions occurring earlier or no interruptions at all should not be as likely to result in PCEs. This is because the main goal of the primary task is still not completed after the interruption and may still propagate activation to the PC subgoal.

So far the CAPS account seems compatible with the interruption position effect on PCE. However, there is one problem without a time-based decay process in the CAPS model because it has difficulty in explaining the occurrences of errors immediately after an interruption. Without a time-based decay mechanism, the CAPS account implies that when the system re-engages with the unaccomplished primary task goal, activation spreads to its subgoals without any diminishing effect. This results in correct resumption every time after an interruption, which is contrary to the results. Furthermore, the CAPS model does not take into account the role of environmental cues and this is mostly because the model was originally developed to investigate a different phenomenon related to PCE, namely working memory demand. The lack of the notion of environmental cues in the account makes it difficult to explain the resumption pattern of PCE, i.e., how one moves on to the next task.

Polson et al.'s supergoal killoff account has a similarity to the CAPS account in that subgoals receive activation and remain active as along as the top-level main goal is unaccomplished and active. The difference in the supergoal killoff account is that it takes into account feedback from the environment such that, "the goal generation process will be driven almost exclusively by prompts and other feedback received from the interface." (p. 745). In this respect, the supergoal killoff account may be extended to explain the interruption position effect on PCE. Regarding interruptions occurring just before the PC step, the account explains the result of PCE immediately after the interruption in terms of the loss of activation supply to the PC subgoal due to the deactivation of the accomplished main goal. Moreover, the account explains the resumption pattern of moving on to the next task in terms of feedback from a false completion signal in the interface. However, the supergoal killoff approach has difficulty in explaining the relatively low likelihood in PCE occurrences after an earlier interruption or no interruption at all. The goal generation process is proposed as solely dependent on feedback from the interface; this predicts that when confronted with the false completion signal, one should only move on to the next task forgetting about the PC step.

There are two further issues with the supergoal killoff account. Firstly, without a memory decay process, it is not clear how the account could explain the generally high occurrences of errors immediately after an interruption. Secondly, the account proposed the occurrence of PCE due to similarity between the PC step and its preceding step. However, as Byrne & Bovair pointed out, the notion of similarity is difficult to operationalise. In terms of the current experimental paradigm, it is hard to argue what should be the criteria to determine similarity between the PC step and its preceding step. For example, should the criteria be on a physical action level (e.g. button click) or on a semantic level of the actions involved?

The Soar account has quite a different theoretical foundation to the other PCE accounts in that it has no notion of activation. The occurrence of PCE is due to the asymmetry between the initiation and the termination of a subgoal. The termination of a subgoal cannot be used to trigger further activities, and when terminated the subgoal and its associated structures disappear. When the main goal is completed and terminated, its associated PC subgoal just disappears, resulting in a PCE. However, the Soar account suggests the presence of compensatory mechanisms, such as self-rehearsal or external cueing in the environment, to explain how the error is mostly avoided.

The account may be stretched to explain that immediately after an interruption, it is more likely for a PCE to occur because the interrupting activity disrupts the compensatory strategies in overcoming the error. As a consequence, there is no rehearsal about the execution of the PC step and it is omitted. On the other hand, the account may explain fewer PCEs occurring with earlier interruptions or no interruptions at all due to the compensatory mechanisms being able to kick in again just before the PC step. However, the Soar account does not use the theoretical construct of activation, and hence no notion of a decay process. Like the other PCE accounts, without a decay process, the Soar approach has difficulties in explaining the generally higher error occurrences immediately after an interruption.

The current discussion suggests that the effect of interruption position on PCE is better explained by the AGM model than the other three PCE accounts. This is hardly surprising as the CAPS account, the supergoal killoff account and the Soar account were not developed to address the effect of interruption on the error. Nevertheless, attempts were made to extend these accounts to see how they might provide insights into the current findings. On a superficial level, each of these accounts seems able to provide a descriptive explanation of the effect of interruption on PCE. However, any account that attempts to explain the effect adequately also needs to explain how an immediately preceding interruption is more likely to result in error than an earlier interruption or no interruption at all. This is because the empirical results suggest that the effect is not unique to PCE but also affects errors in executing the correct task sequence. At this level, the CAPS model, the supergoal killoff account and the Soar account all have problems in offering an adequate explanation. Furthermore, one of the current results suggest that PCE might be more vulnerable to interruption than errors due to remembering ones place in a task sequence. However, none of the existing accounts could explain this finding adequately.

The next chapter is a concluding chapter to summarise the findings and to discuss their implications. The concluding chapter ends with an outline of the contribution to knowledge of the current thesis and suggestions for further work.

Chapter 7

Conclusion

The first goal of the current thesis has been to identify empirically factors that would provoke or mitigate PCE. This goal has been achieved by two series of experiments studying PCE in problem-solving (Chapter 3) and the effects of interruption on PCE (Chapter 6).

Factors provoking PCE

The first series of experiments shows that PCE can be provoked in a novel experimental paradigm — problem-solving — obtaining an error rate above the 5% systematicity criterion. This finding extends the general finding that the occurrence of PCE is sensitive to working memory demands (Byrne & Bovair, 1997; Mortensen, 2003) by requiring participants to solve problem tasks "in the head" without aid from the external environment. The successful generation of PCE in problem-solving tasks shows that anecdotal evidence of PCE occurring in problem-solving activities such as programming (Polson et al., 1992) and simple arithmetic can be studied as empirical phenomena in an experimental setting. This extends our understanding of PCE in that it does not only occur in procedural tasks, which involve executing a set of learned and pre-defined procedures, but also in problem-solving tasks, where task execution involves consciously "working out" the solutions. Thus, it demonstrates the pervasiveness and robustness of PCE as a cognitive phenomenon.

The use of a problem-solving paradigm also constitutes a methodological contribution to the study of PCE. The problem-solving paradigm differs from the procedural task paradigm used in previous studies (e.g. Byrne & Bovair, 1997) in two ways. First, the demand on working memory is an integral component to the problem-solving task itself, and this does not require the use of a dual task paradigm, as in Byrne and Bovair's study, where participants are required to carry out a concurrent secondary task. Secondly, the generation of PCE does not require extensive training on the participants' part in doing the problem tasks. Future experiments based on this problem-solving paradigm can be designed to address questions that are not answered by the current experiments. One of the limits of the current experimental design is that it does not examine the effect of problem complexity on PCE. The problem tasks used in the current experiments are not isomorphic and, consequently, differ in their complexity. Future research can use isomorphic problems or determine problem complexities according to some objective metrics. A hypothesis to be tested is that problems with different complexities might impose different working memory demands, and this might generate different PCE rates.

The hypotheses concerning the effects of interruption position and duration on PCE were tested out in the second series of experiments using a procedural task paradigm. The AGM model (Altmann & Trafton, 2002) provides a theoretical tool to link two research areas together, namely interruption and PCE. The main finding of interruption position suggests that occurrence of PCE is sensitive to where an interruption occurs in the task: a PCE is more likely to occur immediately after an interruption than after an interruption occurring earlier in the task. This finding is consistent with the AGM model's prediction that the increase of PCE immediately after an interruption is because of the disruption of associative cueing from the previous task step and decay of the suspended goal. However, it would be misleading to conclude that the effect of interruption position is unique to PCE. Findings of other errors suggest that errors in remembering the correct step in the task sequence (referred to as incorrect-sequence error) are also sensitive to an immediately preceding interruption.

The hypothesis about the effect of interruption duration on PCE was not supported by the current findings. It was found that PCE is sensitive to an interruption duration as brief as 15 seconds. However, analysis on other errors suggests that incorrect-sequence error is less likely to occur with the 15-second interruption when compared to a 45-second interruption. This finding implies that while PCE and other errors, such as incorrect-sequence error, are sensitive to the disruptive effect of interruption, PCE seems relatively more vulnerable to the short interruption. Studies on interruption duration (e.g. Monk & Trafton, 2004) show that very short interruptions can impose cognitive overheads on primary task performance. Future research can adopt even shorter interruption duration to examine if the disruptive effect on PCE would disappear.

Factor mitigating PCE

The effect of static visual cue was investigated in the series of problem-solving experiments. It was found that a simple static cue implemented in pop-up menus can reliably reduce the number of PCE occurrences. Although the effect of static visual cues is limited in that they do not completely eliminate PCE, they are not completely ineffective in combating PCE as claimed by previous research (Chung & Byrne, 2004). A practical implication of this finding suggests that when it is impractical to implement a dynamic cue, which can occur just-before the PC step, because a system cannot determine the end of an interaction, for example in VCR programming, a static cue can still be a remedy to reduce PCE.

Interpretation of the findings suggests that the complexity of an interface should be taken into account in determining the effectiveness of a visual cue. While static visual cues might not be effective in mitigating PCE when used in interfaces with relatively complex interactions and controls, such as the Phaser task (Byrne & Bovair, 1997; Chung & Byrne, 2004), simple static cues can be effective in combating PCE in simple interfaces involving menu-style interactions. A limitation of this finding is that the notion of interface complexity is only loosely based on contrast between the menu-style interface, in Experiments 3b & 3c, and the interface of Byrne and Bovairs' Phaser task. A future research direction to pursue is what constitutes a complex or visually distracting interface, and invoking psychological theories of attention or visual attention might be a possible starting place to generate more specific hypotheses to determine the complexity of an interface in more specific terms. A further direction to pursue the effectiveness of static visual cue is to investigate its benefit with extend time period. A speculation on this issue is that the mitigating effect of static visual cues on PCE might diminish in the long run. This is because participants might get habituated to the effect of static cues in a prolonged period as suggested by Lee's study (1992).

Relations of PCE in problem-solving and procedural tasks

The empirical investigations of the current thesis suggest that PCE can occur in problem-solving and procedural tasks and this raises the question of to what extent these tasks relate to each other in relation to PCE. There are two aspects to this question: first, it is reasonable to suggest that the cognitive processes involve in the two task types are quite different. In problem-solving activities, no pre-existing knowledge specifying how to solve the task is present and the problem-solver has to "work out" the solution within the constraints of the problem. On the other hand, pre-existing knowledge of how to execute various steps in a procedural task is present and task execution becomes a matter of retrieving the correct information from memory. Although the cognitive processes that underlie problem-solving and procedural task execution might be different, it does not necessarily mean there should be different mechanisms responsible for PCE occurrences in the two types of task. Secondly, the essence of the PC step in the two task types is that its execution is not knowledge-dependent; in other words, to carry out the PC step does not require one to "work it out" but to remember to retrieve such action from memory. Viewing PCE from this perspective, it might be a possible future research direction to invoke prospective memory research to see if it aids the understanding of PCE. Although traditional prospective memory research (e.g. Kvavilashvili & Ellis, 1996) might not regard PCE as a prospective memory problem, the fact that a PC step constitutes a functionally isolated, or not logically inherent, step to the main task might indicate a requirement of prospective remembering in order to carry it out.

Theoretical insights into PCE

The second goal of the current thesis is to gain theoretical insights into PCE and this is achieved by the meta-theoretical analysis (Chapter 4), which critically examines the existing accounts of PCE. The main problem with the existing accounts of PCE is that they have quite different theoretical constructs and, as a result, are difficult to compare and contrast against each other. The critical analysis is able to bring the different extant theoretical accounts under a coherent structure by addressing three fundamental aspects of the error: (1) how the task in question is carried out, (2) how PCE arise from the task performance behaviour, and (3) how the infrequent but persistent nature of the error is manifested.

One of the outcomes of the meta-theoretical analysis is the identification of associative cueing as an important element in the execution of procedural tasks. Among the existing accounts, only the AGM model (Altmann & Trafton, 2002) incorporates associative cueing in proposing how a procedural task is carried out, as the meta-theoretical analysis argues that this is the foundational level, which an account of PCE in procedural tasks should be built upon. The notion of associative cueing between procedural task steps allows predictions to be made with respect to the effect of interruption position on PCE based on the AGM model. Moreover, extending the notion of associative cueing to problem-solving tasks, it is argued that such cueing is minimal and this might also contribute to the high PCE rate generated in the problem-solving paradigm (Experiment

3b). Although the AGM model provides a verbal description of how a procedural task is carried out and motivates the study of interruption on PCE, it does not specify the mechanistic details of *how* associative cueing occurs. The notion of associative cueing provides a good starting point in composing a model of PCE, and there are outstanding questions regarding how sequential behaviour occurs in procedural tasks, such as how do different task steps link to each other? Is a task step only linked to its immediate neighbours or does it also link with task steps in proximity? If so, how strong are the links? These questions provide a basis for future research looking into sequential behaviour in procedural tasks, which itself is a substantial research area.

The meta-theoretical analysis suggests that any adequate account of PCE should be able to account for its infrequent but persistent nature. To address this, it is suggested that a model should specify how PCE is avoided by some compensatory strategies in order to exhibit the infrequent aspect. At the same time, the compensatory strategies should be fragile enough to allow the persistent aspect to occur under cognitively adverse conditions. The implication of this notion of infrequent but persistent nature might be more general than a characteristic unique to PCE, it might apply to many cases of human error that can be classified as slips. The analysis also identifies the occurrence of a false completion signal as an important characteristic of PCE, and, when examined closely, not all the existing accounts, apart from the CAPS model (Byrne & Bovair, 1997) and Reason's (2002) analysis, include this characteristic in their description of PCE. Future research can examine whether it is possible to vary the strength of a false completion signal and investigate if this would have an effect on PCE.

Although it is beyond the scope of the current thesis to propose a complete unified theoretical account of PCE, the meta-theoretical analysis identifies some important characteristics contributing to the occurrence of PCE.

Relation of PCE to other errors

Analyses of the results from Experiments 6b, 6c and 6d also looked at the effects of interruption on other errors as well as PCE. This raises the issue of what is the relation of PCE to other errors; more specifically, is PCE special?

In relation to the effects of interruption, it would be misleading to conclude that PCE is unique because consistent results from Experiments 6b, 6c and 6d suggest that other errors, such as errors involving remembering the correct task sequence (incorrect-sequence error), are also sensitive to the disruptive effect of interruption. The effect of interruption on PCE and incorrect-sequence error suggests that information held in working memory, which allows the correct task step to be executed at the right time, has a dynamic nature and is easily disrupted by interruptions. However, results from the effect of interruption duration suggest that while PCE is sensitive to interruption as brief as 15 seconds, incorrect-sequence error is less affected by the 15-second interruption. This finding suggests that although PCE might not be unique in response to the effect of interruption, PCE might be more vulnerable to the disruptiveness of interruption when compared to other errors. This might be related to a distinguishing feature of PCE, namely the presence of a false completion signal. Upon resumption to the primary task, the false completion signal might act as a miscue cueing one to move on to the next task and omit the PC step.

In contrast to PCE and incorrect-sequence error, it was also found that there are other errors, such as skip-selector error (which involves remembering the correct task sequence but forgetting the correct usage of the interface), that do not seem to be affected by interruptions. The occurrence of skip-selector error is particularly pronounced at the very first step (DP selector) of the doughnut-making task. This error is highly similar to the USW error (Lee, 1992) that it is the omission of a secondary step at the beginning of a procedure. A future experiment can examine the effect of prolonged practice on the error. According to Lee's finding, if the DP skip-selector error is similar to USW error, then the error rate should decrease with initial practice and increase again with more practice. An interesting manifestation of the DP skip-selector error is that its occurrence is high at the first step of the doughnut-making task and that it is not due to misremembering the correct sequence of task itself. The occurrence of the DP skip-selector error suggests that the dynamic information, for place-keeping, held in working memory is intact but the knowledge of using the interface correctly is forgotten. This provides a contrast to PCE, whose occurrence is due to disruption to dynamic information in working memory rather than knowledge of using the interface correctly.

Explaining two real-world PCE examples

The following is a description of two PCE cases and the purpose of this is to bring the general findings of the current thesis together by articulating how PCE is likely to arise in each case.

Case 1: The voice message system at a certain university requires the caller to press the # key after saying the message in order to save the message in the system.

In the above case, it is predicted that there will be a high proportion of people forgetting to press the # key after speaking the message, and consequently, the message is not recorded. The prediction is based on the following rationales: first, it is not unreasonable to say that most voice message systems do not require the extra step of pressing the # key at the end of a message, therefore, this extra step is probably not in most users' procedural knowledge when leaving a voice message. In other words, the procedure of finishing speaking the message is unlikely to act as an associative cue to prime the execution of pressing the # key: associative cueing is minimal, if at all, between these two steps. Secondly, when leaving a voice message there is a considerable demand on the user's working memory assuming that the user is not reading out loud from a script but has to construct what he/she has to say. Thirdly, when the user finishes speaking the message this gives a strong false completion signal indicating the end of the task. The combination of a lack of associative cueing, high working memory demand and appearance of a false completion signal is likely to give rise to PCE in the above case — forgetting to press the # key after leaving the message.

Case 2: A user of a photocopier is making a copy of a single-sided document. When the copy is made and collected, the user is interrupted by his/her boss with a 5-minute conversation just before he/she is going to collect the original document.

In this case of a photocopying task, it is predicted that the user is likely to make a PCE — forgetting to collect the original document — due to the interruption. In contrast to the case 1, the procedure of the PC action — collecting the original — is in the user's procedural knowledge assuming that the user is an experienced user of the photocopier.

Therefore, under usual circumstances, the action of collecting the desired copy should act as an associative cue priming the action of collecting the original. However, the conversation with the user's boss interrupts the associative cueing process, and the length of the interruption is likely to be long enough for the subgoal of collecting the original to decay below retrieval threshold. When the conversation is over, the user resumes to the photocopying task seeing the desired copy in his/her hands, and this acts as a false completion signal indicating the end of the task. Moreover, there is no visual cueing about the PC step because the original document is hidden under the lid of the photocopier. Since the subgoal of collecting the original has decayed during the length of the conversation, seeing the desired copy in his/her hand the user leaves and forgets about the original document.

The above two cases are used as examples to illustrate the general findings of PCE from the current thesis. Although it is possible that there are other factors that might have contributed to the occurrence of PCE in the above cases, such as the user's motivation, importance of the original document, time pressure, etc., it is not to claim that the above cases are explained exhaustively by the current findings. The use of the PCE cases are examples to show that we can anticipate more reliably about the occurrences of PCE in certain situations than we previously know about the error phenomenon.

Contribution to knowledge

The current thesis contributes to the substantive knowledge of PCE through a series of empirical investigations and a critical theoretical analysis. The empirical findings show that, first, the occurrence of PCE extend beyond procedural tasks to problem-solving tasks; secondly, PCE is sensitive to where an interruption occurs in a primary task and that an interruption as brief as 15 seconds can increase the occurrence of PCE. A simple static visual cue is found to mitigate the occurrence of PCE reliably, although complete elimination of the error is not possible.

Critical examination of existing PCE accounts through the meta-theoretical analysis identifies some important criteria that any adequate model of PCE should take into account in future theoretical development. Although it is beyond the analysis to achieve a complete account of PCE, it is the first attempt to compare and contrast the different theoretical accounts within a coherent structure.

Finally, the use of a novel experimental paradigm in Chapter 3 — problem-solving — and another procedural task in Chapter 6, in addition to the Phaser task (Byrne & Bovair, 1997), also makes a methodological contribution to the study of PCE. The very first experiment in each of the two series of experiments (Chapter 3 and 6) has met difficulty in generating PCE, but various methodological shortcomings were identified and subsequently rectified. These tasks not only contribute to the limited task repertoire in studying PCE, but also meet one of the main difficulties in studying human error in laboratory settings: to generate a high enough error rate for subsequent examination.

References

Adamczyk, P.D., & Bailey, B. P. (2004). If not now, when? The effects of interruption at different moments within task execution. *Proceedings of CHI'04* (pp. 271-278). New York: ACM Press.

Allwood, C. M. (1984). Error detection processes in statistical problem solving. *Cognitive Science, 8*, 413-437.

Altmann, E. M. (2003). Task switching and the pied homunculus: where are we being led? *Trends in Cognitive Science, 7(8)*, 340-341.

Altmann, E. M. (2004). Advance preparation in task switching. *Psychological Science, 15(9)*, 616-622.

Altmann, E. M. & Trafton, J. G. (1999). Memory for goals: an architectural perspective. *In Proceedings of the 21st Annual Conference of the Cognitive Science Society*

Altmann, E. M., & Trafton J. G. (2002). Memory for goals: an activation-based model. *Cognitive Science, 26*, 39-83.

Altmann, E. M., & Trafton, J. G. (2004). Task interruption: resumption lag and the role of cues. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society: Erlbaum.*

Anderson, J. R., & Liebere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Baars, B. J. (1992). The many uses of error: Twelve steps to a unified framework. In B. J. Baars (Ed.) *Experimental slips and human error: Exploring the architecture of volition.* New York: Plenum Press.

Baddeley, A. D. (1990). *Human Memory: theory and practice.* LEA.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2000). Measuring the Effects of Interruptions on Task Performance in the User Interface. *IEEE International Conference on Systems, Man, and Cybernetics 2000, 2*, 757-762.

Bainbridge, L. (1984). Diagnostic skills in process operation. *Proceedings of 1984 International Conference on Occupational Ergonomics, 2*, 1-10.

Ben-Zeev, T. (1995). The nature and origin of rational errors in arithmetic thinking: Induction from examples and prior knowledge. *Cognitive Science, 19*, 341-376.

Burmistrov, I., & Leonova, A. (1996). Effects of Interruptions on the Computerised Clerical Task Performance. In S. Overmyer & J. Gornostaev (Eds.), *Human-Computer Interaction: Human Aspects of Business Computing*, 21-29. Moscow: ICSTI.

Byrne, M. D. & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive Science, 21*, 31-61

Byrne, M. D. & Davis, E. M. (in press). Task structure and postcompletion error in the execution of a routine procedure. To appear in *Human Factors.*

Byrne, M. D. (2003). A mechanism-based framework for predicting routine procedural errors. *In Proceedings of the 25th Annual Conference of the Cognitive Science Society.*

Card, S., Moran, T. & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Erlbaum.

Chung, P. H. & Byrne, M. D. (2004). Visual cues to reduce errors in a routine procedural task. *In Proceedings of the 26th Annual Conference of the Cognitive Science Society.*

Cutrell, E.B., Czerwinski, M., & Horvitz, E. (2000). Effects of Instant Messaging Interruptions on Computing Tasks. *Proceedings of CHI2000* (pp. 99-100). New York: ACM Press.

Curzon, P. & Blandford, A. (2001). Detecting multiple classes of user errors. In Proceedings of the 8th IFIP Working Conference on Engineering for Human Computer Interaction. LNCS. Springer-Verlag.

Curzon, P. (2003). "The dog, the hen and the corn...and post completion errors". *Interfaces, 55*, 19-22, British HCI Group.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant messaging: Effects of relevance and time. In S. Turner & P. Turner (Eds.), *People and Computers XIV: Proceedings of HCI 2000, 2,* 71-76.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. *Proceedings of CHI'04* (pp. 175-182). New York: ACM Press.

Degani, A, & Wiener E. L. (1990). *Human Factors of Flight-Deck Checklists: The normal Checklist.* (NASA-CR-177549). Washington, DC: National Aviation and Space Administration.

Dismukes, K., Young, G., & Sumwalt, R. (1998). Cockpit Interruptions and Distractions. *ASRS Directline*, 10.

Dix, A. J., Finlay, J. E., Abowd, G. D., & Beale, R. (1998). *Human-Computer Interaction (Second Edition).* Prentice Hall Europe.

Edwards, M. B., & Gronlund, S. D. (1998). Task interruption and its effects on memory. *Memory, 6(6)*, 665-687.

Einstein, G. O., McDaniel, M. A., Williford, C. L., Pagan, J. L., & Dismukes, R. K. (2003). Forgetting of intentions in demanding situations is rapid. *Journal of Experimental Psychology: Applied, 9(3)*, 147-162.

Ernst, G. W. & Newell, A. (1969). *GPS: A case study in generality in problem solving.* New York: Academic Press.

Eyrolle, H., & Cellier, J.-M. (2000). The effects of interruptions in working activity: field and laboratory results. *Applied Ergonomics, 31*, 537-543.

Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.

Gamble, T. R., & May, J. (2002). Increasing trial foreperiod in task switching. *Poster presented at European Working Memory Symposium I*, Ghent, Sept. 5-7, 2002.

Gamble, T. R., & May, J. (2003). Task Switching and Switch Cost between task sequences. *Poster presented at BPS Cognitive Section Conference, Reading, Sept 2003. Proceedings of the British Psychological Society, 12*, 104.

Gillie, T., & Broadbent, D. (1989). What makes interruption disruptive? A study of length, similarity, and complexity. *Psychological Research, 50*, 243-250.

Gray, W. D. (2000). The nature and processing of errors in interactive behaviour. *Cognitive Science, 24(2)*, 205-248.

Gray, W. D. (2004)[10]. Errors in interactive behavior . In W. S. Bainbridge (Ed.), *Encyclopedia of Human-Computer Interaction* (pp. 230-235): Berkshire Publishing Group. *Also available at http://www.rpi.edu/~grayw/pubs/papers/wdg04 HCI-Encycl.html (last visited: Sept. 2006)*

Greeno, J. G. (1974). Hobbitts and Orcs: Acquisition of a sequential concept. *Cognitive Psychology, 6(2)*, 270-292.

Goschke, T., & Kuhl, J. (1993). Representation of intentions: Persisting activation in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19(5)*, 1211–1226.

Hodgetts, H. M., & Jones, D. M. (2003). Interruptions in the Tower of London task: can preparation minimise disruption? *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society.*

Hodgetts, H. M., & Jones, D. M. (2005). Interrupting problem solving: effects of interruption position and complexity. *Proceedings of the 40th Australian Psychological Society Annual Conference*, Melbourne, Australia (pp. 128-132).

Hodgetts, H., M., & Jones, D., M. (2006a). Interruption of the Tower of London task: support for a goal-activation approach. *Journal of Experimental Psychology: General, 135(1)*, 103-115.

[10] Page numbers cited from Gray (2004) in this thesis are based on the downloaded copy from the corresponding URL.

Hodgetts, H., M., & Jones, D., M. (2006b). Contextual cues aid recovery from interruption: the role of associative activation. *Journal of Experimental Psychology: Learning, Memory & Cognition.* Manuscript in press.

Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method.* Oxford: Elsevier.
Jeffries, Polson, Razran & Atwood's (1977)

Jeffries, R. P., Polson, P. G., Razran, L. & Atwood, M. (1977). A process model for missionaries-cannibals and other river-crossing problems. *Cognitive Psychology, 9*, 412-440.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 95*, 163-182.

Kieras,D. E., & Polson, P. G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies, 22*, 365-394.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Kitajima, M. & Polson, P. (1992). A computational model of skilled use of a graphical user interface. *In CHI '92 Conference Proceedings.* Reading, MA: Addison Wesley.

Kotovsky, K. & Kushmerick, N. (1991). Processing constraints and problem difficulty in problem-solving: a model. *In Proceedings of the 13th Annual Conference of the Cognitive Science Society.*

Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*, 248-294.

Kvavilashvili, L. & Ellis, J. A. (1996). Varieties of intention: Some distinctions and classifications. In M. Brandimonte, G. O. Einstein, M. A. McDaniel (Eds.) *Prosepective memory: Theory and Applications.* Erlbaum: Mahwah, N.J.

Laird, J. E., Newell, A. & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*, 1-64.

Larson, G. E., & Perry, Z. A. (1999). Visual capture and human error. *Applied Cognitive Psychology, 13*, 227 – 236.

Latorella, K. A. (1999). Investigating interruptions: implications for flightdeck performance. *NASA/TM-1999-209707*, National Aeronautics and Space Administration, Hampton.

Lee, W. O. (1992). The effects of skill development and feedback on action slips. In A. Monk, D. Diaper and M. Harrison (Eds.) *People and Computers VII.* Cambridge: Cambridge University Press.

Lee, W. O. (1993). Incremental change in the development of expertise in using interactive systems. Unpublished doctoral thesis. University of Cambridge.

McDaniel, M. A., Einstein, G. O., Graham, T., & Rall, E. (2004). Delaying execution of intentions: overcoming the costs of interruptions. *Applied Cognitive Psychology, 18*, 533-547.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt.

Monk, A. (1986). Mode errors: a user-centred analysis and some preventative measures using keying-contingent sound. *International Journal of Man-Machine Studies, 24*, 313-327.

Monk, C. A. (2004). The effects of frequent versus infrequent interruptions on primary task resumption. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society.*

Monk, C, A., & Trafton, J. G. (2004). Very brief interruptions result in resumption cost. *Poster presented at the 26th Annual Meeting of the Cognitive Science Society.*

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at carious stages. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1824 – 1828). Santa Monica, CA: HFES.

Monk, C, A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors, 46(4)*, 650 - 663.

Monsell, S. (2003). Task switching. *Trends in Cognitive Science, 7(3),* 134-140.

Mortenson, I. C. (2003). *An investigation of working memory areas: post-completion errors and the implications for HCI.* Unpublished Masters Thesis: Middlessex University.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88*, 1-15.

Norman, D. A. (1983). Design rules based on analyses of human error. *Communications of the ACM, 26(4)*, 254-258.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books.

Oulasvirta, A., & Saariluoma, P. (2004). Long-term working memory and interrupting messages in human-computer interaction. *Behaviour and Information Technology, 23(1)*, 53-64.

Ovisankina, M. (1928) Die Wiederaufnahme unterbrochener Handlungen. *Psychologische Forschung, 11 (1)*, 302-379.

Payne, S. J. & Squibb, H. R. (1990). Algebra mal-rules and cognitive accounts of errors. *Cognitive Science, 14*, 445-481.

Polson, P.G., Lewis, C., Rieman, J. & Wharton, C. (1992). Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *International journal of Man-Machine Studies, 36*, 741-773.

Rasmussen, J. (1980). What can be learned from human error reports? In K. Duncan, M. Gruneberg & D. Wallis (Eds.) *Changes in Working Life.* London: Wiley.

Rasmussen, J. (1982). A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4*, 311-333.

Rasmussen, J. (1987). Cognitive control and human error mechanisms. In J. Rasmussen, K. Duncan & J. Leplat (Eds.) *New technology and human error.* London: John Wiley & Sons.

Rasmussen, J., Duncan, K. & Leplat, J. (1987). *New technology and human error.* London: John Wiley & Sons.

Rasmussen, Pejtersen & Goodstein, (1994) Cognitive Systems Engineering. John Wiley & Sons: New York, NY.

Ratwani, R. M. & Trafton, J. G. (2006). Now, where was I? Examining the perceptual processes while resuming an interrupted task. *Proceedings of the 28th Annual Conference of the Cognitive Science Society.*

Ratwani, R. M., Trafton, J. G. & Myers, C.(2006). Helpful or harmful? Examining the effects of interruptions on task performance. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society.*

Reason, J. (1990). *Human Error.* New York: Cambridge University Press.

Reason, J. (2002). Combating omission errors through task analysis and good reminders. *Quality of Safety and Health Care, 11*, 40-44

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124(2)*, 207-231.

Rouse, W. & Rouse, S. (1983). Analysis and classification of human error. *IEEE Transactions on Systems, Man, and Cybernetics, 13(4)*, 539-549.

Sanders, M. S. & McCormick, E. J. (1993). *Human Factors in Engineering and Design.* Singapore: McGraw-Hill.

Schoppek, W. (2000) Leanring and performance of sequential action under different workload conditions: an ACT-R model. *In Proceedings of the 3rd International Conferences on Cognitive Modelling.* Veenendaal: Universal Press.

Sellen, A. J. & Norman, D. A. (1992). The psychology of slips. In B. J. Baars (Ed.) *Experimental slips and human error: Exploring the architecture of volition.* New York: Plenum Press.

Senders, J. W. & Moray, N. P. (1991). *Human error: Cause, prediction and reduction.* Hillsdale, NJ: Erlbaum.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7(2)*, 268-288.

Speier, C., Vessey, I., & Valacich, J. S. (2003). The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decision Sciences, 34(4)*, 771-797.

Thimbleby, H. (1990). *User Interface Design.* ACM Press.

Thomas, J. C. (1974). An analysis of behaviour in the hobbits-orcs problem. *Cognitive Psychology, 6(2)*, 257-269.

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58*, 583-603.

Van Bergen, A. (1968). *Task Interruption.* Amsterdam: North-Holland.

Walji, M., Brixey, J., Johnson-Throop, K., & Zhang, J. (2004). A theoretical framework to understand and engineer persuasive interruptions. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.*

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Context.* London: Batsford.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch cost". *Psychological Research, 63*, 212-233.

Young, R. M. (1994). The unselected window scenario: Analysis based on the Soar cognitive architecture. In *CHI '94 Research Symposium*, 9-11.

Young, R. M., Green, T. R. G. & Simon, T. (1989). Programmable user models for predictive evaluation of interface designs. *ACM SIGCHI Bulletin, 20*, 15-19.

Zacks, J., & Tversky, B. (2001). Event structure in perception and cognition. *Psychological Bulletin, 127(1)*, 3-21.

Zeigarnik, B. (1927). Das Behalten erledigter und unerledigter Handlungen. *Psychologische Forschung, 9*, 1-85.

Zijlstra, F. R. H., Roe, R. A., Leonora A. B., & Krediet, I. (1999). Temporal factors in mental work: effects of interrupted activities. *Journal of Occupational and Organizational Psychology, 72*, 163-185.

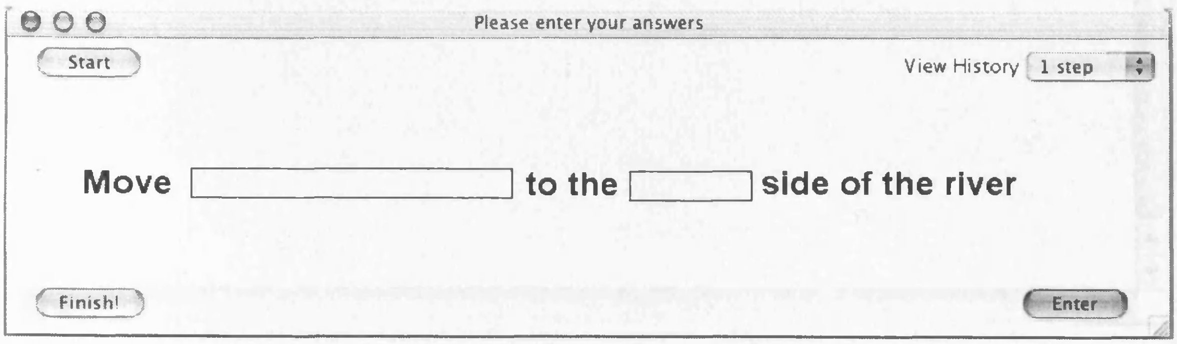
Appendix A (i)

Interface instruction manual

Thank you for participating in this Human-Computer Interaction (HCI) study.

In this study, you will be asked to solve a series of problems. Each problem will be described to you on a problem sheet which contains all the necessary information to solve the problem. Please inform the experimenter once you have finished reading the problem sheet and it will be taken away. You will then be asked to recall the locations and the objects that you think you will have to enter into a computer in solving the problem. For example, if one of the objects in a problem is a person named John you should recall the name "John" and not just "a person".

Each problem involves a series of steps to be solved. You will be required to enter the answers into a computer. The following is how you would enter your answers through the specific interface:



After recalling the locations and the objects to be manipulated, please press the "*Start*" button immediately.

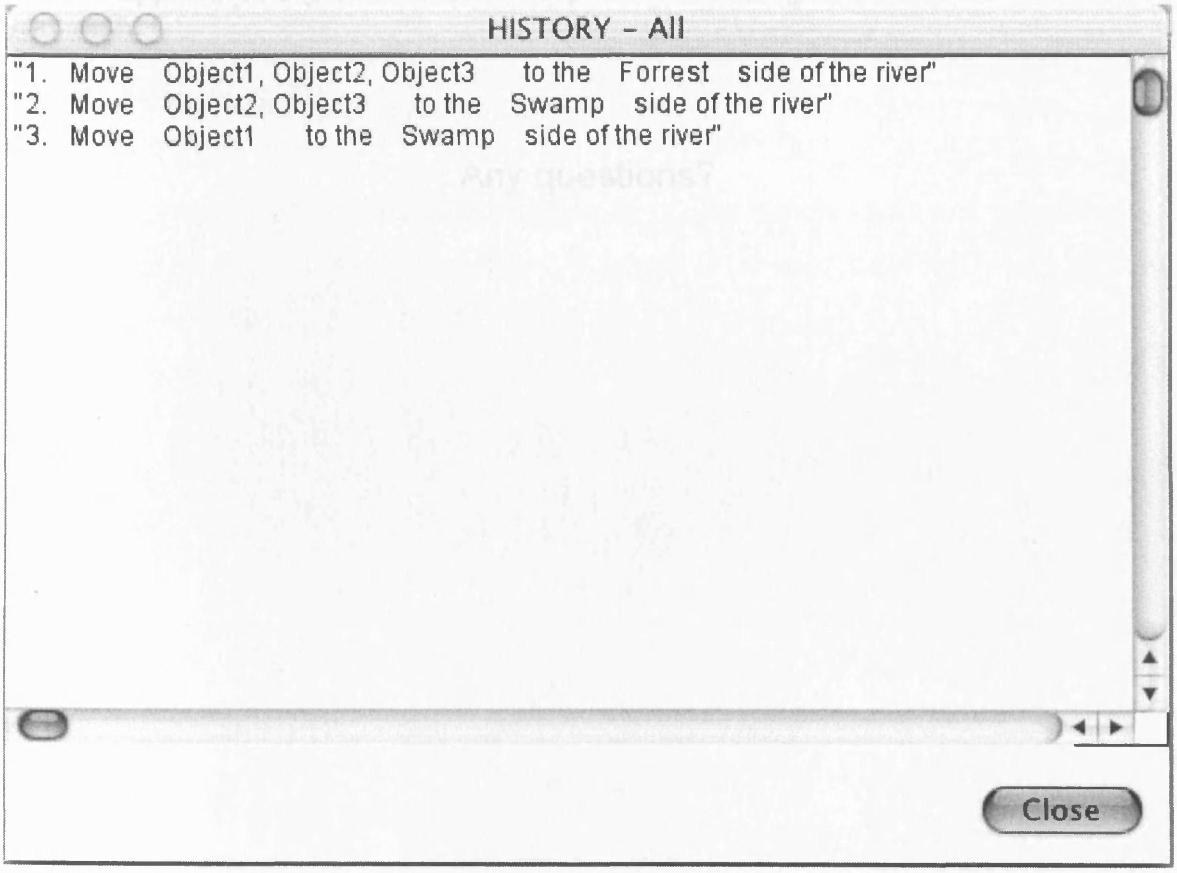
Type your answers into the text boxes through the keyboard. You can move between the text boxes by hitting the Tab key or via the mouse. Once you are happy with the answer you have typed in, you need to hit the "*Enter*" button by simply pressing the Return key on the keyboard or by using the mouse to click the "*Enter*" button on the screen.

You are allowed to enter more than one object into the long text box on the left. However, please separate the objects with a comma and a space when you do so. For example, if you want to enter Peter, Apple and Toy then you should type in:

```
Peter, Apple, Toy
```

When you think you have reached the solution click on the "Finish" button.

You can use the "*View History*" drop-down menu to look back at what you did. You will see three options when you click on the menu. " *1 Step*" will display the previous answer you gave; "2 *Steps*" will display two previous answers you entered; and "*All*" will allow you to look at all the previous answers you gave. The most recent answer is always at the bottom of the window. When you have finished looking at what you did before, you can return to entering your answers by clicking the "*Close*" button:



The "View History" option might come in handy when you've forgotten what you did previously, however, please **ONLY** use this option when you feel it is absolutely necessary. Furthermore, you should also use the three options strategically, i.e. do not view all previous answers when you only need to look at the two previous steps.

You will have the opportunity to interact with the interface in a test trial before the main trial.

Your performance will be timed and timing will begin from the point you start reading the problem sheet. It is very important that you should perform as **accurately** and **quickly** as possible.

You will be asked to do the same thing for all the problems. In between problems, you are allowed to take a 3-min break. Remember this is **NOT** a test of your IQ so please do not feel pressured.

Any information you provide in this study will be kept confidential. The study will last for approximately 30 minutes and you have the right to leave the session if you wish.

Any questions?

Appendix A (ii)

Training problem sheet

In this problem you will need to get Object39, Object72 and Object58 from the Swamp side of a river to the Forest side. Since this is only a test trial, the sequence of steps to solve the problem are already provided and you just need to type in the information accordingly. This is to get you comfortable with the interface that you will be using.

Please memorise the objects and the sides of the river as you will be asked to write them down on a separate sheet of paper. You may begin this test trial by handing this sheet back to the experimenter.

Appendix A (iii)

Recall sheet

Objects:

Sides of the river:

Appendix A (iv)

Training problem instruction sheet

Step 1: Start trial by clicking on the "Start" button

Step 2: Send Object39, Object72, Object58 to the Forest side of the river

Step 3: Send Object39, Object58 to the Swamp side of the river

Step 4: View previous step made by selecting "1 step" from the drop-down menu of "View History"

Step 5: Send Object58 to the Forest side of the river

Step 6: View 2 previous steps made by selecting "2 step" from the drop-down menu of "View History"

Step 7: Send Object72, Object58 to the Swamp side of the river

Step 8: View all previous steps made by selecting "All" from the drop-down menu of "View History"

Step 9: End the trial by clicking on the "Finish" button

Appendix A(v)

Reminder sheet

In the series of problems which you will be asked to solve, they all involve moving objects from one side of the river to the other. Anything that you move across the river are classified as objects. For example a nurse, Fred, a cat, a carrot and a canoe are all classified as objects.

There are different rules which you have to follow in order to solve each problem in a particular sequence of steps. Each step you make, you have to indicate all the necessary objects that will be moved across the river by entering them into the left-hand-side text box and the corresponding side of the river into the right-hand-side text box.

Appendix A (vi)

The Father & Son (FS) problem

Bob and his two sons, Jim and Chris, want to go to the Christmas fair near their village. In order to get to the fair, which is on a small island next to the village, they have to go across a river. The only means of conveyance is a small boat which is moored on the river bank. The maximum capacity of the boat is 200 pounds and any weight more than that will sink it. The boat does not have a auto-pilot system and cannot get across the river without a passenger.

Jim and Chris both weigh 100 pounds each while Bob weighs 200 pounds. Assuming all three of them can operate the boat, how can they all get from the village side to the fair side of the river without getting anyone soaking wet?

Appendix A (vii)

The Dog, Hen & Corn (DHC) Problem

Mrs. Jones, a farmer, is on her way to the market with her trusty dog, Shep. The market is on the other side of a river. Getting to market always used to involve a long detour round the river. However, Mr. Edison, the local inventor who lives at the bank of the river decided to rig up a rope-pulley contraption to allow them to get straight across. It consists of a rope slung between pulleys on either bank, with a seat just big enough for one person plus one thing to carry with. Whoever finishes using the seat always return it to the market side, where Mr. Edison lives, so that it is easy for him to take it in each evening. On arriving at the river when going to the market Mrs. Jones pulls the seat across from the far side using the rope. She gets in, hugging Shep then pulls herself across the river and continues into the market.

On the occasion in question Mrs. Jones buys a new hen and a sack of corn from the market. Returning home, from the market side, later in the day she arrives back at the river, and quickly realises she has a problem. She can only carry one thing across with her at a time and she has to be on the seat in order to offload whatever she is bringing with her when reaches the other side of the river. She will have to make more than one trip. This time it is worse than that however: if she leaves the hen and the corn alone on either side, the hen will eat the corn; similarly if she leaves Shep and the hen together on one side Shep will worry the hen, which may mean it stops laying eggs. Shep is not interested in eating corn so it will come to no harm with him.

How can Mrs. Jones get everything across, from the market to the home side of the river, uneaten and continue on her way?

Appendix A (viii)

The simplified Missionaries & Cannibals (sMC) Problem

In the island — Tau — the missionaries have been attacked by the cannibals and both parties are now fighting for survival. After a few days of battle, there remains two missionaries and two cannibals. While the cannibals never say no to more dinner and the missionaries are fighting to save their lives, both parties spot a boat by the left side of a river in the island. The boat is connected to a rope-pulley system and the passenger has to send the boat back to where it was, in order for other passengers to get across, after it has reached the opposite side of the river. Of course there is no other way to get across the river apart from using the boat!

Intrigued by the discovery the two cannibals wanted to explore the other side of the island and the missionaries are hoping to be able to escape from the cannibals if they could get to the other side of the river. With the mutual objective of reaching the other side of the river, both parties agree to stop the battle and try crossing the river on the boat.

Although all four of them know how to operate the boat, it has a maximum capacity of only one person. This means that the passenger has to disembark before the next trip. To make things worse, if, at any time, there are more cannibals than missionaries on either side of the river, those missionaries will be eaten by the cannibals!

How can all four get from the left side to the right side of the river without any missionaries being eaten?

Appendix A (ix)

Solutions for the FS problem:

	Solution 1		Solution 2	
Step	**Object(s)**	**Side of the river**	**Object(s)**	**Side of the river**
1	Chris, Jim, boat	fair	Chris, Jim, boat	fair
2	Chris, boat	village	Jim, boat	village
3	Bob, boat	fair	Bob, boat	fair
4	Jim, boat	village	Chris, boat	village
5	Chris, Jim, boat	fair	Chris, Jim, boat	fair

Solutions for the DHC problem:

	Solution 1		Solution 2	
Step	**Object(s)**	**Side of the river**	**Object(s)**	**Side of the river**
1	Mrs. Jones, hen, seat	home	Mrs. Jones, hen, seat	home
2	Mrs. Jones, seat	market	Mrs. Jones, seat	market
3	Mrs. Jones, Shep, seat	home	Mrs. Jones, corn, seat	home
4	Mrs. Jones, hen, seat	market	Mrs. Jones, hen, seat	market
5	Mrs. Jones, corn, seat	home	Mrs. Jones, Shep, seat	home
6	Mrs. Jones, seat	market	Mrs. Jones, seat	market
7	Mrs. Jones, hen, seat	home	Mrs. Jones, hen, seat	home
8	seat	market	seat	market

Appendix A (ix)

Solution for the sMC problem:

	Solution 1	
Step	**Object(s)**	**Side of the river**
1	Cannibal, boat	right
2	boat	left
3	Missionary, boat	right
4	boat	left
5	Missionary, boat	right
6	boat	left
7	Cannibal, boat	right

Appendix B (i)

Training problem sheet

Toby is an architect and he has a passion for assembling models of not just buildings but also toy cars, tanks, planes etc. He lives in a house, which is designed by himself, just outside London. His house is in a very quiet spot and is separated from the nearest town by a very charming river. Across the river is Toby's favourite coffee shop, whenever he fancies a coffee (and maybe a piece of cake as well sometimes) he would have to make a detour and drive around the river to get to this coffee shop. One day Toby has this rather neat idea of building his own little remote-controlled toy boat to bring coffee across the river to his house so that he doesn't have to drive! The little boat has two purposely-shaped containers to carry a coffee and a piece of cake. It cannot carry two coffees or two cakes at the same time. Now the ordering involves just a simple phone call to the coffee shop to place the order and the staff will then put the coffee (and cake sometimes) into the little boat for Toby to steer back to his house. After each order, the boat is always sent back to the coffee shop side of the river, as a convenience to the staff for the next order.

Today, Toby has a couple of friends around to his house for coffee and he decided to show off his new toy! The total order came to a Latte, a Cappuccino, a filter coffee, a muffin and a brownie. How can Toby get all of the order back to his house from the coffee shop with a minimum number of trips across the river? Please also send the boat back to the coffee shop when finished taking everything across.

Appendix B (ii)

Check list

The following statements are related, one way or the other, to the story you just read, some of them are true while some are not. Please tick next to the statements which you think are true. If the statement were true, please indicate how important (Not at all, Moderate or Very) it would be to you in solving the problem.

☐ Toby is an architect

Not at all　　Moderate　　Very

☐ The total order includes a Cappuccino, a Latte, a filter coffee and a muffin

Not at all　　Moderate　　Very

☐ The boat is built by Toby's friends

Not at all　　Moderate　　Very

☐ The total order includes a Cappuccino, a Latte, a filter coffee, a brownie and a muffin

Not at all　　Moderate　　Very

☐ The boat can carry two cakes

Not at all　　Moderate　　Very

☐ Toby lives opposite a coffee shop across a river

Not at all　　Moderate　　Very

☐ The boat is always sent back to the house side of the river after use

Not at all　　Moderate　　Very

☐ Toby lives opposite a toy shop across a river

Not at all Moderate Very

☐ The boat can carry two coffees

Not at all Moderate Very

☐ The boat can carry one coffee and one cake

Not at all Moderate Very

☐ Toby is an interior designer

Not at all Moderate Very

☐ The boat is always sent back to the coffee shop side of the river after use

Not at all Moderate Very

☐ The boat is built by Toby

Not at all Moderate Very

☐ The total order includes a Cappuccino, a filter coffee, a brownie and a muffin

Not at all Moderate Very

Appendix B (iii)

The Lion, Monkey & Banana (LMB) Problem

Jerry is a circus trainer and one day he is going back to the circus after some shopping in town. This time Jerry is going back with his favourite pet, Rambo, a well-tamed lion, a monkey and a banana. The circus and the town are separated by a river and the only way to get across it is by using a boat, which is big enough for Jerry to carry two things with him.

Crossing the river is a little tricky for Jerry this time. If he leaves the monkey and the banana alone on either side of the river, the cheeky monkey will eat the banana before they get back to the circus; similarly if he leaves Rambo and the monkey alone on either side Rambo will harm the monkey. But Rambo is not interested in the banana so it will come to no harm with him. Jerry has to be in the boat in order to offload whatever he is bringing with him when getting across the river.

How should Jerry get everything across the river, from the town side to the circus side, uneaten and unharmed?

Appendix B (iv)

The 3 Guests (3G) Problem

Mario and Luigi live in a small village where things are simple and quiet. The village is separated from the main town by a river. To get into town, Mario and Luigi would use a boat to get across the river.

One day, Mario and Luigi had three guests over for a dinner party. After the lovely dinner, Mario and Luigi had to escort their guests to the town side of the river. Getting across the river needs a bit of planning this time, since the boat is only big enough for Mario and Luigi together or one of their guests at a time. The boat is not even big enough for one guest with Mario or Luigi. So how can all the guests get across to the town side of the river?

Appendix B (v)

The Torch (Trch) Problem

Adam, Ben, Charles and David all go to the same secondary school and are very good friends with each other. The four of them always hang out together and sometimes they are a bit adventurous and spend a night in a forest nearby. This is considered a daring night out for them as they do this without their parents' permission! And of course they will all be in trouble if found out as the forest is not a particularly safe place to be in at night. One night the four of them are out in the forest again for their adventurous exploration. This time they find a river which they had not noticed before. On the bank there is a boat attached to a pulley system which allows it to be sent across the river. Being all adventurous and daring, the boys decide to get across the river and find out what lies on the other side! When they reach the other side of the river, they have to send the boat back to the forest so that no one will notice it has been used.

However, getting across the river is not a straight forward task. It is totally dark and the boys are sharing one torch between them. The light torch only has batteries to last for 17 mins. Each person takes different times to get across the river; Adam 10 mins, Ben 5 mins, Charles 2 mins and David 1 min. None of them is brave enough to get across without the torch. The boat can only take a maximum of two people at a time and it will only go at the slower person's speed. How can all four get from the forest to the other side of the river in 17 mins? The torch cannot be transported on its own using the boat.

Appendix B (vi)

The Itchy & Scratchy (IS) Problem

Two mafia gangs, Itchy and Scratchy, in New York have planned a bank robbery operation in Lower Manhattan. The operation involved three members from each family. In this targeted bank, it is not just the cash they are interested in but also a suitcase of rare black diamonds. Both gangs got the money and also the suitcase of diamonds. As they went out of the bank, they were surrounded by the FBIs! However, both gangs did not just surrender but began a street war with the FBIs. After a series of gun shots and car chase, the two gangs managed to lose the FBIs and went outside the city. They reached the Hudson river and there they found a reasonably sized crate attached to a pulley system which the locals use for transporting goods across the river. Both parties decided to go to the other side of the river which they thought as a temporarily safe hiding place. They realise that they have to send the crate back to the initial side when they reach the other side so no one would suspect that it had been used.

Each gang wants the suitcase of diamonds for themselves and the dynamics of the two gangs have now changed... The Itchies have used up all their bullets and the Scratchies will shoot the Itchies whenever there are more Scratchies than Itchies on either side of the river. Neither gang trusts the other with the suitcase, this means that the suitcase cannot be left on either side of the river while crossing. The suitcase needs to be in the crate all the time with a gang member in case the crate tips and it falls into the river. It cannot be transported on its own to the other side. The crate is only big enough for two people max. How can all members of the two gangs go from the initial side to the other side of the river without anyone getting shot?

Note: If there is one Scratchy on one side already and you are moving one Itchy and another Scratchy to that same side, this will cause the Itchy to be shot even if one of the Scratchies immediately go to the other side. So such a move is an illegal move.

Appendix B (vii)

Solution for the training problem:

	Solution	
Step	**Object(s)**	**Side of the river**
1	Muffin, Filter Coffee	House
2	Boat	Coffee Shop
3	Latte, Brownie	House
4	Boat	Coffee Shop
5	Cappucino	House
6	Boat	Coffee Shop

Note: The choice of coffee and cake to be transported does not have to be those as indicated. The combination of coffee and cake is not fixed as long as it conforms to the constraint of the problem.

Appendix B (viii)

Solution for the LMB problem:

	Solution	
Step	Object(s)	Side of the river
1	Jerry, Rambo, Banana	Circus
2	Jerry	Town
3	Jerry, Monkey	Circus

Appendix B (ix)

Solution for the 3G problem:

	Solution 1		Solution 2	
Step	Object(s)	Side of the river	Object(s)	Side of the river
1	Mario, Luigi	Town	Mario, Luigi	Town
2	Luigi	Village	Mario	Village
3	Guest	Town	Guest	Town
4	Mario	Village	Luigi	Village
5	Mario, Luigi	Town	Mario, Luigi	Town
6	Luigi	Village	Mario	Village
7	Guest	Town	Guest	Town
8	Mario	Village	Luigi	Village
9	Mario, Luigi	Town	Mario, Luigi	Town
10	Luigi	Village	Mario	Village
11	Guest	Town	Guest	Town
12	Mario	Village	Luigi	Village

Appendix B (x)

Solution for the Trch problem:

	Solution 1		Solution 2	
Step	**Object(s)**	**Side of the river**	**Object(s)**	**Side of the river**
1	Charles, David	Other	Charles, David	Other
2	David	Forest	Charles	Forest
3	Adam, Ben	Other	Adam, Ben	Other
4	Charles	Forest	David	Forest
5	Charles, David	Other	Charles, David	Other
6	Boat	Forest	Boat	Forest

Appendix B (xi)

Solution for the IS problem:

	Solution 1		Solution 2	
Step	**Object(s)**	**Side of the river**	**Object(s)**	**Side of the river**
1	Itchy, Scratchy	Other	Scratchy, Scratchy	Other
2	Itchy	Initial	Scratchy	Initial
3	Scratchy, Scratchy	Other	Scratchy, Scratchy	Other
4	Scratchy	Initial	Scratchy	Initial
5	Itchy, Itchy	Other	Itchy, Itchy	Other
6	Itchy, Scratchy	Initial	Itchy, Scratchy	Initial
7	Itchy, Itchy	Other	Itchy, Itchy	Other
8	Scratchy	Initial	Scratchy	Initial
9	Scratchy, Scratchy	Other	Scratchy, Scratchy	Other
10	Scratchy	Initial	Scratchy	Initial
11	Scratchy, Scratchy	Other	Scratchy, Scratchy	Other
12	Crate	Initial	Crate	Initial

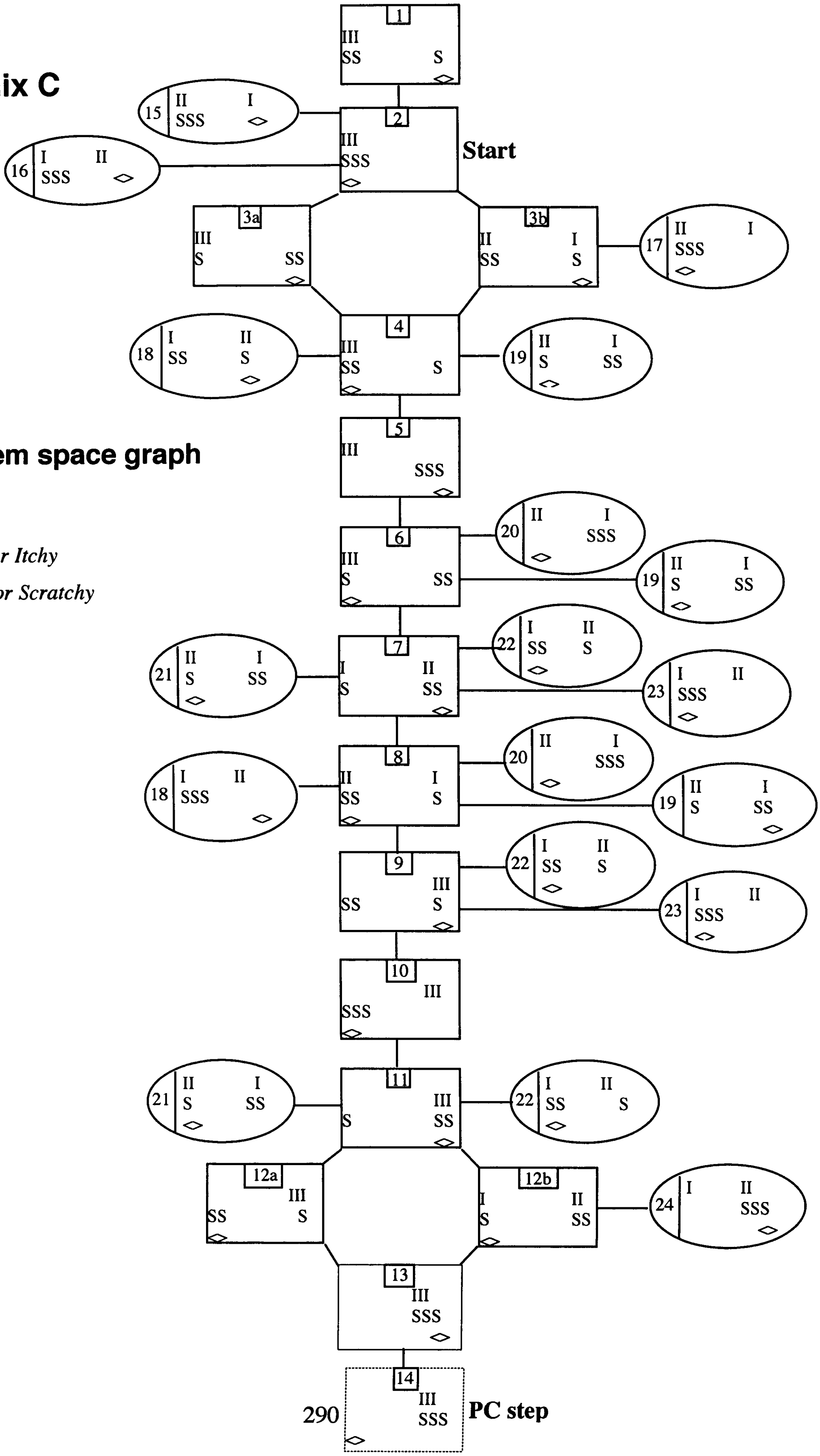
Appendix C

IS problem space graph

Note:

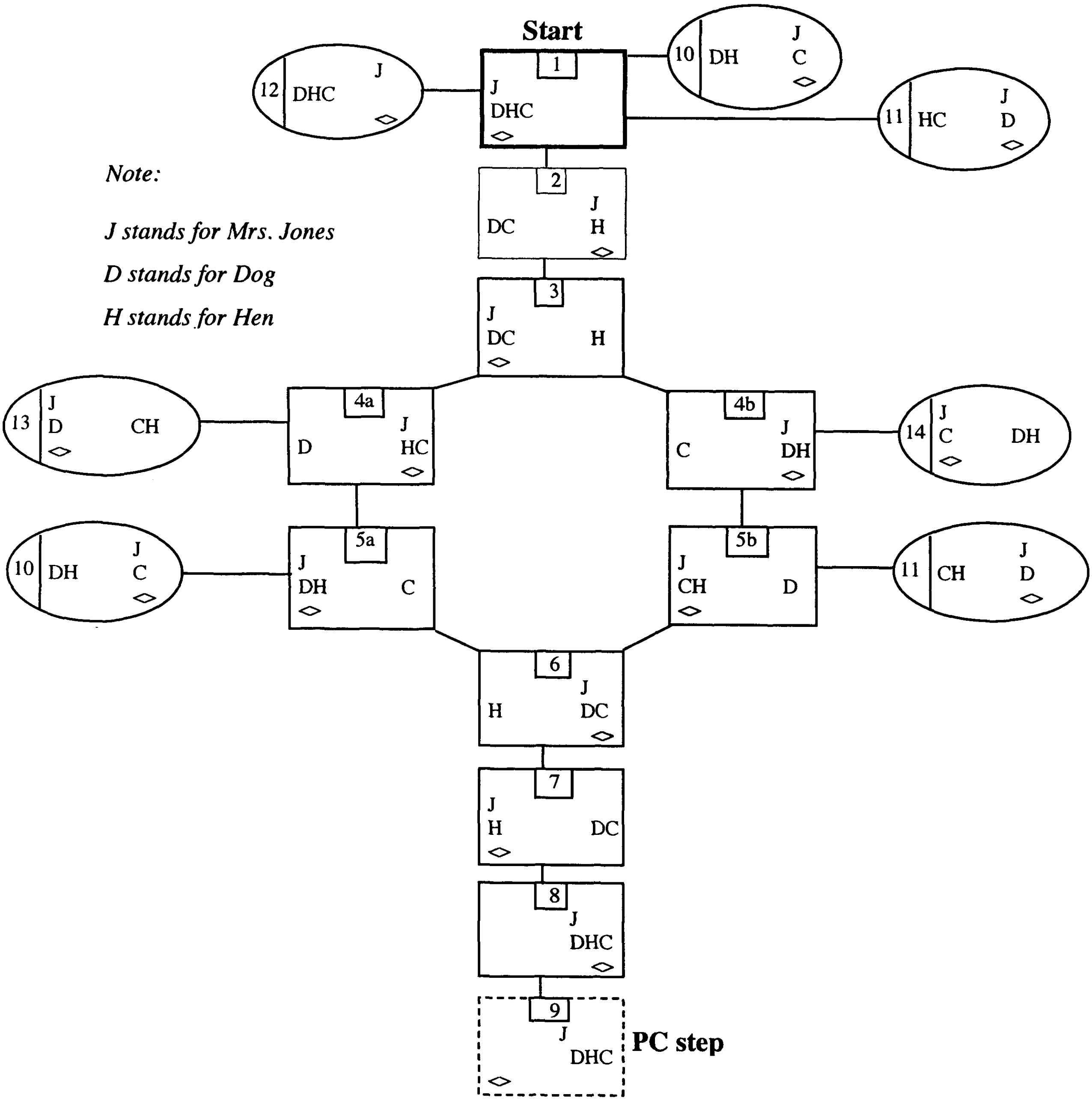
I stands for Itchy

S stands for Scratchy



Appendix C

DHC problem space graph



Note:

J stands for Mrs. Jones

D stands for Dog

H stands for Hen

Appendix C

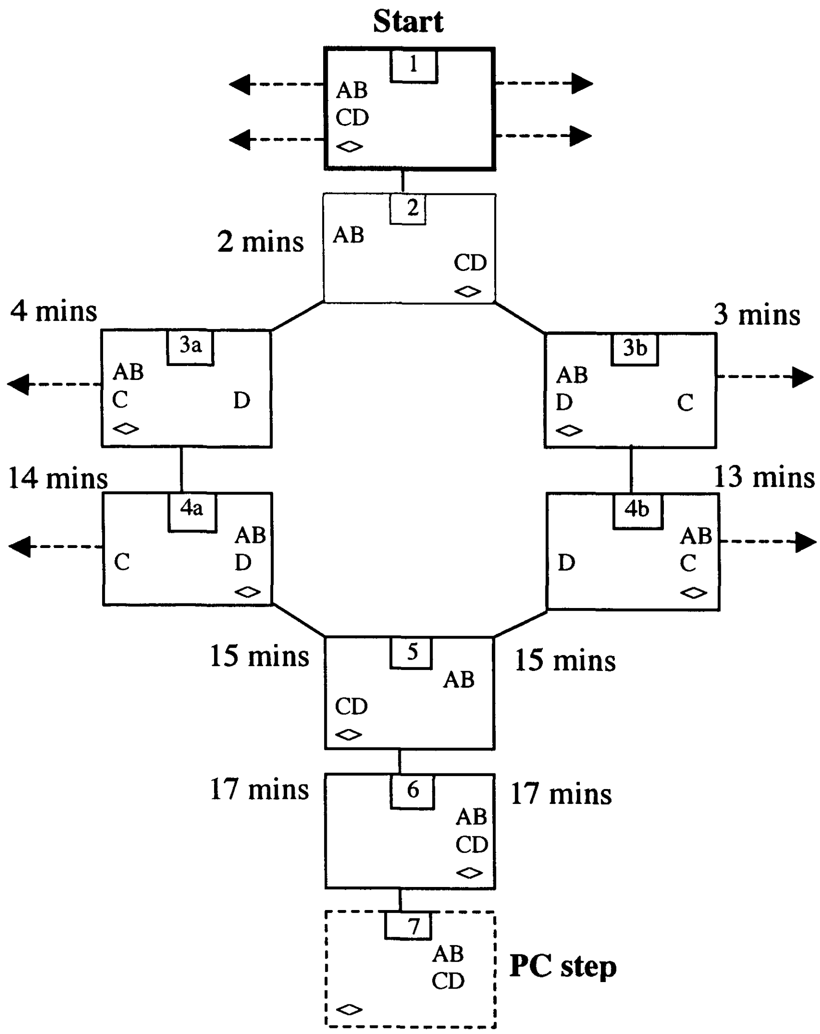
Trch problem space graph

Note:

A stands for Adam
B stands for Ben
C stands for Charles
D stands for David

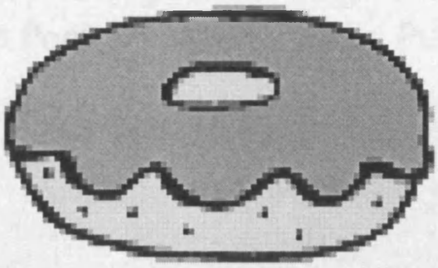
Note:

A dotted arrow indicates a path can lead to an illegal state with too many branches to be drawn out



Appendix D

Thank you for taking part in this study of Human-Computer Interaction. In this session you will be playing a computer-based game called "*The Wicket Doughnut Machine*".



In this game you are the baker and you have two tasks:

1) Make doughnuts using the Wicket Doughnut Machine. However, the doughnut machine is slightly faulty, and you, as the baker, has to try your best to make the right amount of doughnuts ordered! It's no good to your bakery to produce more than the order and it's definitely not a good way to keep your customers to make less than the order!

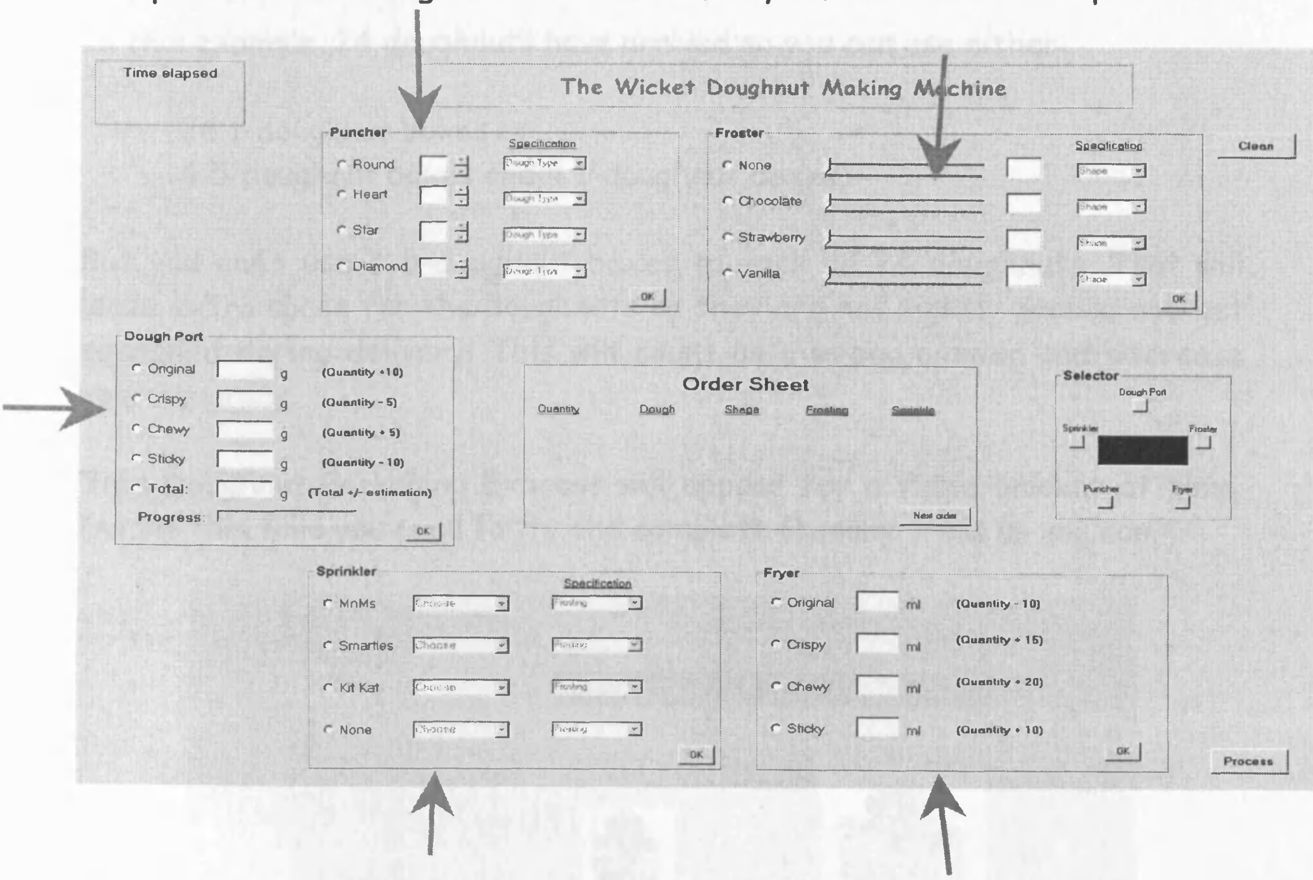
2) Pack doughnuts into boxes via the Doughnut Packaging Express

There is a specific way to operate the machine and you will be walked through the instructions step-by-step.

Appendix D

The Wicket Doughnut Machine

Below is the Wicket Doughnut Machine… it has five main compartments: Dough Port, Froster, Fryer, Puncher and Sprinkler.



The five compartments need to be operated in the following sequence:

Dough Port ⑧ Puncher ⑧ Froster ⑧ Sprinkler ⑧ Fryer

Otherwise, the machine will not work!

Appendix D

Doughnut Packaging Express

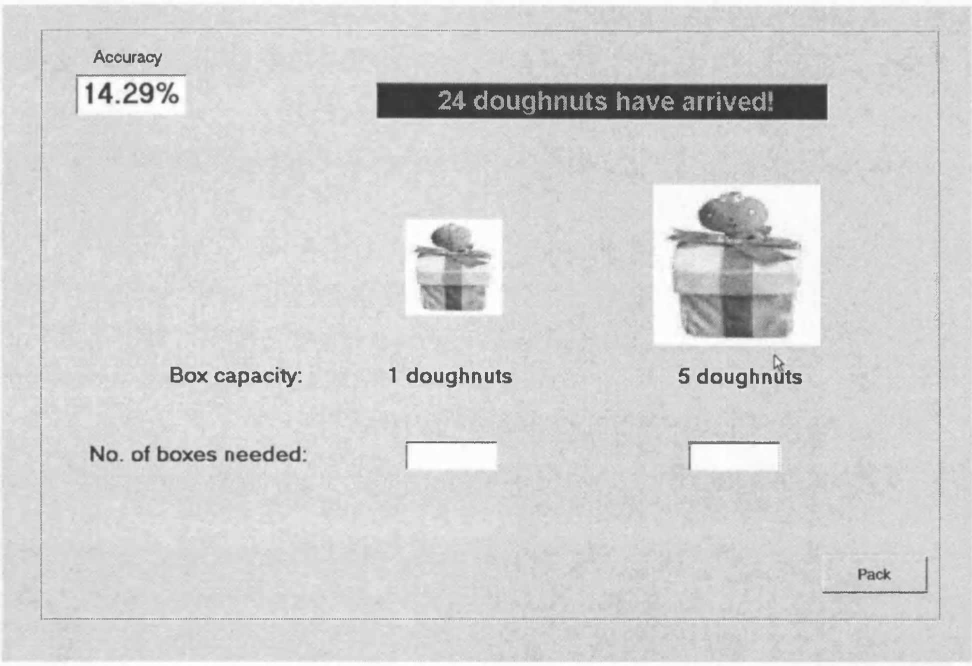
During some testing trials you may see a screen, as below, while operating the doughnut machine. When such screen appears you need to pack the number of arrived doughnuts into two different sizes of boxes.

In this example, 24 doughnuts have arrived so you can use either:

> 24 1-doughnut boxes or,
> 4 5-doughnut boxes and 4 1-doughnut boxes.

But you can't use 5 5-doughnut boxes to pack all 24 doughnuts. That will leave extra space for the doughnuts so they are not tightly packed and get squashed during delivery! This will count as a wrong answer and decrease your accuracy.

This Doughnut Packaging Express will appear for a fixed amount of time. During this time you need to try and complete as many trials as you can.



Appendix D

You will need to do a total of 10 trials on the Wicket Doughnut Machine. After the first 5 trials you will be asked if you would like to have a break and may do so as you wish.

This is not an assessment or a test of IQ. You are only required to follow the instructions just presented and carry out the trials. However, you are encouraged to perform as quickly and accurately as you can in each task and in each trial.

Any questions?

Appendix E

Fitts' Law calculations for Experiment 6a

The purpose of these calculations is to provide an approximation of the physical movement times involved in resuming the PC step and the various "Selector" step. For both sets of calculation, the starting point of the resumption movement is assumed to be the location of the "Pack" button in the interrupting Doughnut Packing task. The calculations are only meant to be rough approximations, therefore, various estimations that have also been made are stated.

Fitts' Law states that:

Movement time (in msec) $= a + b \log_2(D / S + 1)$

where a and b are parameters determined empirically. D is the distance of the movement measured and S is the size of the target (usually the width of a button). For the current purpose the values of a and b used are adopted from Raskin's (2000) recommendation, which are 50 and 150 respectively.

Calculation for resumption movement time of the PC step:

S = 2.3 cm (width of the "Clean" button)

D = 16.5 cm (this is the distance measured from the centre of the "Pack" button, on the interrupting task screen, to the centre of the "Clean" button on the primary task screen, which is at the upper right corner of the screen layout)

Movement time (in msec) $= 50 + 150 \log_2(16.5 / 2.3 + 1)$
$= 505$ msec

Calculation for resumption movement time of a "Selector" button:

S = 0.4cm (width of a "Selector" button)

D = 6.5 cm (rather than measuring distances of all 5 "Selector" buttons individually, the distance is measured from the centre of the "Pack" button, on the interrupting task screen, to the space in the middle of the 5 "Selector" buttons on the primary task screen. This measurement should serve as a "good enough" approximation as the 5 buttons are relatively close in proximity.)

Movement time (in msec) $= 50 + 150 \log_2(6.5 / 0.4 + 1)$
$= 670$ msec

Fitts' Law calculations for Experiment 6b, 6c and 6d

The approach of the calculations for Experiment 6b, 6c and 6d is the same as 6a.

Calculation for resumption movement time of the PC step:

S = 1.9cm (width of the "Clean" button)

D = 4.5 cm (this is the distance measured from the centre of the "Pack" button, on the interrupting task screen, to the centre of the "Clean" button on the primary task screen, which is at the lower right corner of the screen)

Movement time (in msec) = $50 + 150 \log_2(4.5 / 1.9 + 1)$
= 310 msec

Calculation for resumption movement time of a "Selector" button:

This is the same calculation as Experiment 3.1 since there were no changes made to the position or size of the "Selector" buttons and the "Pack" button in Experiment 6b, 6c and 6d.